



University of Kentucky
UKnowledge

Theses and Dissertations--Educational, School,
and Counseling Psychology

Educational, School, and Counseling
Psychology

2015

EFFECTIVENESS OF PSYCHOTHERAPY FOR YOUTH IN POVERTY: A BENCHMARKING STUDY OF A PUBLIC BEHAVIORAL HEALTH AGENCY USING A CLIENT FEEDBACK SYSTEM

Jonathan Kodet

University of Kentucky, jkodet@gmail.com

Digital Object Identifier: <http://dx.doi.org/10.13023/ETD.2016.015>

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Kodet, Jonathan, "EFFECTIVENESS OF PSYCHOTHERAPY FOR YOUTH IN POVERTY: A BENCHMARKING STUDY OF A PUBLIC BEHAVIORAL HEALTH AGENCY USING A CLIENT FEEDBACK SYSTEM" (2015).

Theses and Dissertations--Educational, School, and Counseling Psychology. 41.

https://uknowledge.uky.edu/edp_etds/41

This Doctoral Dissertation is brought to you for free and open access by the Educational, School, and Counseling Psychology at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Educational, School, and Counseling Psychology by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Jonathan Kodet, Student

Dr. Robert Jeffrey Reese, Major Professor

Dr. Kenneth Maurice Tyler, Director of Graduate Studies

EFFECTIVENESS OF PSYCHOTHERAPY FOR YOUTH IN POVERTY: A
BENCHMARKING STUDY OF A PUBLIC BEHAVIORAL HEALTH AGENCY
USING A CLIENT FEEDBACK SYSTEM

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Education
at the University of Kentucky

By

Jonathan David Kodet

Lexington, Kentucky

Director: Dr. Robert Jeffrey Reese, Professor of Counseling Psychology

Lexington, Kentucky

2015

Copyright © Jonathan David Kodet 2015

ABSTRACT OF DISSERTATION

EFFECTIVENESS OF PSYCHOTHERAPY FOR YOUTH IN POVERTY: A BENCHMARKING STUDY OF A PUBLIC BEHAVIORAL HEALTH AGENCY USING A CLIENT FEEDBACK SYSTEM

Objective: The purpose of this study was to evaluate the effectiveness of a public behavioral health (PBH) agency that had implemented continuous outcome feedback as a quality improvement strategy.

Method: I investigated the pre-post treatment outcomes of 4,389 ethnically diverse youths (6 to 17 years old) at or under the poverty line participating in treatment (from January 2008 to March 2014) for a broad range of primary diagnoses including depression and anxiety disorders (23%); adjustment disorders (27%); Attention Deficit/Hyperactivity Disorder (13%), various V-codes (18%); bipolar disorders (3%); and substance use disorders (2%). I also investigated the treatment outcomes for a subset of youth ($N = 469$) presenting with depression-related psychological distress. Treatment outcome was measured with the Outcome Rating Scale (ORS; Duncan, 2011; Miller & Duncan, 2004) and the child version: Child Outcome Rating Scale (Duncan Sparks, Miller, Bohanske, & Claud, 2006). Benchmark methodology allowed effect size comparisons to randomized clinical trials.

Results: The average treatment effect size estimate of psychotherapy ($d = 0.74$) for all youth at the PBH agency was comparable to the average effect size estimate for treatment from nine clinical trials using client feedback, yet not equivalent to an average effect size estimate from feedback trials using the ORS. Compared to treatment-as-usual (TAU) groups, treatment at PBH was clinically superior to the TAU group outcomes in both the benchmark from all nine feedback trials and the TAU benchmark from the three ORS trials. The average treatment effect size estimate of psychotherapy ($d = 1.51$) for the PBH depression sample was clinically superior to a waitlist/no treatment benchmark drawn from 17 clinical trials of youth depression, and clinically equivalent to a treatment benchmark drawn from 13 youth depression clinical trials using intent-to-treat analyses.

Conclusions: Despite the existing socioeconomic disparities in mental healthcare for youth, these findings demonstrate that mental health services to youth in poverty across an entire agency can be effective. Continuous outcome feedback can bridge the gap

between research and practice and may be a feasible strategy to ensure quality of services for PBH agencies.

KEYWORDS: Benchmarking, Depression, Poverty, Youth, Psychotherapy

Jonathan David Kodet
Student

December 15, 2015
Date

EFFECTIVENESS OF PSYCHOTHERAPY FOR YOUTH IN POVERTY: A
BENCHMARKING STUDY OF A PUBLIC BEHAVIORAL HEALTH AGENCY
USING A CLIENT FEEDBACK SYSTEM

By

Jonathan David Kodet

Robert Jeffrey Reese, Ph.D.
Director of Dissertation

Kenneth Maurice Tyler, Ph.D.
Director of Graduate Studies

December 15, 2015
Date

ACKNOWLEDGEMENTS

My life as a graduate student has been connected to individuals and events that have shaped me forever. I received so much support from many friends, family, mentors, colleagues, and research collaborators. First, I would like to express deep gratitude to my dissertation chair, Jeff Reese, PhD; your phenomenal mentoring provided a constant source of support, inspiration, and motivation throughout each stage of this dissertation and PhD process. I am grateful for your patience and believing in me. I would also like to thank my committee members and outside reader, Rory Remer, PhD., Kenneth Tyler, PhD, Chizimuzo (Zim) T.C. Okoli, PhD., and Jean Wiese, PhD (outside reader). Each of you provided highly personalized and timely support to help me complete this journey. Also, this project would not have been possible without the collaboration of Southwest Behavioral Health Services (SBHS). Special thanks to Bob Bohanske, PhD. from SBHS, in particular, and to Barry Duncan, PsyD. From Heart and Soul of Change Project for making the time to support me and this project.

In addition to mentors and research collaborators, I received exceptional support from my family and many friends. I want to thank my entire family from the bottom of my heart for being supportive of me throughout my graduate education. Special thanks to my three sisters for encouraging me the whole way through, and especially my oldest sister, Julie Gohman, PhD.; your words of wisdom at key moments were crucial to my completion of this project.

Finally, I want to both thank and dedicate this dissertation to the two most important people in my life, my wife and my daughter. Amy and Sage, you bring so much joy and purpose into my life, words do not suffice.

To my daughter Sage, I am so proud to be your dad. Thank you for being so understanding for all the missed events and my days away writing this “story.”

I have been a graduate student now for most of your life. You are a constant source of personal pride and your wisdom, creativity, kindness, and natural presence tell me that there is nothing you cannot accomplish in life. I believe in you. Never forget who you are and where you come from.

To my wife Amy, none of this would have been possible without you. You made so many sacrifices that allowed me to complete this journey and your support in the big and small matters was vital to me achieving this goal. You believed in me, encouraged me, and were there for me when I needed it the most.

My journey to this Ph.D. has not been easy, and I know that I could not have done this alone. I sincerely thank all of the wonderful people, named and un-named, who have made this dissertation possible.

TABLE OF CONTENTS

Acknowledgements	iii
List of Tables	vii
Chapter One: Introduction and Review of Selected Literature	1
Mental Health and Youth in Poverty	3
Psychotherapy With Youth	5
Psychotherapy With Youth in Poverty	7
Benchmarking studies for youth in poverty.	9
Weersing and Weisz (2002).....	10
Curtis, Ronan, Heiblum, and Crellin (2009).....	12
Lee, Horvath, and Hunsley (2013).....	13
Limitations.	14
Summary of problem one.	15
Measuring Progress and Alliance for Quality Improvement	15
Client feedback.	16
Client feedback studies with youth.	18
Client feedback and evidence-based practice.	19
Client feedback and the working alliance.	20
The Partners for Change Outcome Management System (PCOMS).....	22
Summary of problem two.	24
Purpose of This Study	25
Chapter Two: Method	26
Design.....	26
Procedures	27
Participants	28
Measures.....	30
Outcome Rating Scale (ORS).....	30
Child Outcome Rating Scale (CORS)	33
Benchmark Methodology	33
Benchmarks construction.	35
Depression benchmarks.	35
Clinical trial selection.....	35
Use of existing meta-analysis.....	37
Client feedback (complete sample) benchmarks.	42
Depression efficacy trial benchmark effect size calculations.	43
Critical value calculation.	45
Data Analysis	46
Effect size calculations.	46
Benchmarking analyses.	47
Benchmarking against treatment groups.....	47
Benchmarking against waitlist control and treatment-as-usual conditions.....	48
Research Hypotheses.....	48

Hypothesis one.....	49
Hypothesis two.....	49
Hypothesis three.....	50
Hypothesis four.....	50
Chapter Three: Results.....	51
Preliminary Analysis	51
Results of Client Feedback Benchmark Hypotheses.....	55
Hypothesis one.	56
Hypothesis two.	57
Results of Depression Benchmark Hypotheses.....	57
Hypothesis three.	58
Hypothesis four.....	58
Clinical Significance	59
Chapter Four: Discussion.....	61
Effectiveness of Client Feedback with Youth in Poverty	62
Clinical Equivalency to Treatment Benchmarks.....	65
Supporting Previous PCOMS Research.....	66
Study Limitations	66
Implications and Future Recommendations.....	69
Conclusions	72
Appendix A.....	73
Appendix B	74
Appendix C	75
References.....	76
Curriculum Vitae	105

LIST OF TABLES

Table 1, Client Demographic Information for Full and Depression Samples	29
Table 2, Treatment groups for Intent-To-Treat Depression Benchmark	38
Table 3, Control Groups for Waitlist Depression Benchmark.....	40
Table 4, Full Sample Therapy Outcomes by Client Demographics	51
Table 5, Full Sample Therapy Outcomes by Diagnosis	54
Table 6, Depression-related Clinical Sample Therapy Outcomes by Diagnosis	55
Table 7, Effect Size Comparisons to Client Feedback Benchmark RCT Studies	56
Table 8, Effect Size Comparisons to Depression Benchmark RCT Studies.....	58
Table 9, Clinically Significant Change and Reliable Change	60

Chapter One: Introduction and Review of Selected Literature

Mental health problems for youth continue to be a significant challenge for individuals, families, and communities in the United States and globally. The World Health Organization (2012) estimates that up to one in five youth suffer from mental disorders. Specific to the United States, estimates of youth given a diagnosis of a mental disorder range from 13-20% each year accruing an annual cost of almost 247 billion dollars in treatment and related healthcare costs (Centers for Disease Control and Prevention, 2013). Mental health issues are among the most costly of conditions to treat for youth (Soni, 2009). For the first time in the 50 years that the U.S. government has collected data of childhood disabilities, mental health problems now make up the top five disabilities affecting children rather than physical problems (Halfon, Houtrow, Larson, & Newacheck, 2012). A U.S. nationally representative study using principal diagnosis at hospital admittance saw an 80% increase in mood disorders during 1997-2010, from 10 to 17 hospitalizations per 10,000 youth (Pfuntner, Wier, & Stocks 2013). Given the substantial cost to youth, families, and communities, effective psychosocial interventions are urgently needed, and the large-scale evaluation of promising treatment approaches in “real world” clinical settings is a worthy pursuit.

Large-scale effectiveness studies testing evidence-based treatments with youth in public behavioral health (PBH) settings with comparison groups are sorely lacking. Given that a large and potentially growing percentage of youth from economically impoverished backgrounds do not receive adequate mental health care (e.g., Warren, Nelson, Mondragon, Baldwin, & Burlingame, 2010), improving the current state of

psychotherapy in PBH settings would be a financially wise and socially just investment. I will evaluate a client feedback system for that exact purpose.

Client feedback refers to the practice of monitoring client self-report outcome throughout the course of treatment. The research evidence supporting client feedback in psychotherapy is compelling with adults in individual therapy (Lambert, 2013; Miller, Duncan, Brown, Sorrell, & Chalk, 2006; Reese, Norsworthy, & Rowlands, 2009; Whipple et al., 2003), couples therapy (Anker, Duncan, & Sparks, 2009; Reese, Toland, Slone, & Norsworthy, 2010), and group psychotherapy (Schuman, Slone, Reese, & Duncan, 2015; Slone, Reese, Mathews-Duvall, & Kodet, 2015). Yet, few studies have evaluated the benefit of client feedback with youth.

Through the current naturalistic effectiveness study, I will not only evaluate treatment effectiveness by calculating sample pre-post effect size estimates (ESs), but will also employ the most current benchmarking methodology in order to strengthen internal validity (Minami, Serlin, Wampold, Kircher, & Brown, 2008). Benchmarking methodology permits a comparison against established efficacy benchmarks (i.e., ESs from randomized clinical trials of youth diagnosed with depression). Two central questions guided this study. First, in comparison to benchmarks found in efficacy trials, is psychotherapy utilizing a client feedback system effective in reducing psychological distress among youth diagnosed with depression in a PBH setting? Second, is psychotherapy utilizing client feedback effective in reducing psychological distress irrespective of diagnosis among youth in a PBH setting compared to feedback benchmarks? In preparation for answering these questions, the extant literature on

mental health and mental healthcare in relation to youth from economically impoverished backgrounds will be reviewed.

Mental Health and Youth in Poverty

The present day situation in the United States is even bleaker when looking at mental health issues for youth from economically impoverished backgrounds. First, episodic poverty rates (2 months or longer) and chronic poverty rates (at least 36 months) for youth increased from 2005 to 2011 in the United States, and youth had the highest rates of each poverty type (40.6% and 5.9%, respectively, in the most recent 2009-2011 sample) compared to both adult categories of aged 18-64 or 65 years and over (Edwards, 2014). Evans (2004) reviewed research showing the myriad of environmental, relational, and psychological stressors that children and adolescents of low socioeconomic status contend with daily. For example, youth in poverty live in neighborhoods with more crime (in metro areas), street traffic, substandard housing, abandoned lots, boarded up buildings, and inadequate municipal services. Compared to those from higher SES, children from economically impoverished backgrounds have greater noise exposure, are 3.6 times more likely to live in houses infested with rodents, are 2.7 times more likely to have inadequate heat in the winter, and have fewer retail facilities or supermarkets with healthy and discounted foods. Although this list is not exhaustive, it shows how youth of lower SES households and neighborhoods contend with a cumulative effect of multiple stressors that can negatively influence their physical and mental health. The consequences are substantial.

One of the most consistently replicated findings in the social sciences is that for most physical and mental health problems, a SES-health gradient can be seen with worse

outcomes being found at each step down the SES ladder (American Psychological Association, Task Force on Socioeconomic Status, 2007; Frank & Glied, 2006; Muntaner, Eaton, Miech, & O'Campo, 2004). Specific to psychological functioning, the lower the SES of an individual, the higher the risk of mental health problems overall (Dohrenwend, 1990; Fryers, Melzer, & Jenkins, 2003; Hudson, 2005; Jokela, Batty, Vahtera, Elovainio, & Kivimäki, 2013; Pan, Stewart, & Chang, 2013; Pinquart & Sörenson, 2000; Reiss, 2013). Researchers have examined how SES predicts later mental health status with lower SES being a robust predictor of more frequently occurring mental health problems (Bosma, van de Mheen, & Mackenbach, 1999; Costello, Compton, Keeler, & Angold, 2003; Dohrenwend et al., 1992; Gallo, Bogart, Vranceanu, & Matthews, 2005; Hudson, 2005; Kessler & Cleary, 1980; Lundberg, 1997; McLeod & Shanahan, 1993; Ritsher, Warner, Johnson, & Dohrenwend, 2001). This relationship has been stable over time and across different measures of mental health (Frank & Glied, 2006; Lorant et al., 2003) with both adults and youth (McLaughlin, Costello, Leblanc, Sampson, & Kessler, 2012; Merikangas et al., 2010; Strohscchein, 2005). Furthermore, additional studies have shown a strong connection between adversity early in life—with childhood poverty being a main factor—and adult mental health problems (e.g., Case & Paxson, 2006; Cohen, Janicki-Deverts, Chen, & Matthews, 2010; Kessler et al., 2005; Shonkoff, Boyce, & McEwen, 2009). The insidious accumulation of stressors from the adversity of persistent poverty is perhaps the most detrimental to youth mental health (McLeod & Shanahan, 1993). All of this evidence has led to a growing recognition of the importance of intervening early with mental health distress (Karoly, Kilburn, &

Cannon, 2006; Walter et al., 2011) with clinicians, researchers, and policy makers focusing more on improving publically funded mental health services for youth.

Psychotherapy With Youth

Psychotherapy overall has a long history of efficacy through randomized clinical trials with adults (Duncan, Miller, Wampold, & Hubble, 2010; Lambert, 2013) and with youth (Weisz, 2014; Weisz & Jensen, 2001). Efficacy studies on psychotherapy outcomes have shown that clients significantly improve when compared to no treatment, delayed treatment, or being given a placebo (Lambert & Ogles, 2004; Watanabe, Hunot, Omori, Churchill, & Furukawa, 2007). Likewise, early meta-analyses of psychotherapy with youth reported impressive results. Weisz and Jensen (2001) reviewed four meta-analyses of broad-based psychotherapy with children and adolescents. They found effect sizes (ESs) ranging from $d = 0.71$ to $d = 0.84$. According to Cohen's (1988) guidelines for interpreting the magnitude of ESs (ESs between 0.20 and 0.49 [*small*]; ES between 0.50 and 0.79 [*moderate*]; ES of 0.80 or greater [*large*]), ESs of 0.71 and 0.84 are considered to be in the upper end of the moderate range and lower end of the large range, respectively.

Similarly, three early meta-analyses covering psychotherapy for youth depression in clinical trials show a positive picture. In the earliest meta-analysis, Reinecke, Ryan, and DuBois (1998) reviewed six cognitive behavioral therapy (CBT) studies from articles published in peer-reviewed journals and found a mean ES of 1.02. In the second meta-analysis, Lewinsohn and Clarke's (1999) search of published peer-reviewed journal articles resulted in an even larger mean ES of 1.27, on the basis of 12 treatment–control comparison studies. In the third meta-analysis between 1980 and 1999, Michael and

Crowley (2002) included both psychosocial studies and pharmacological trials with youth diagnosed with Major Depressive Disorder. For the 15 controlled trials of psychosocial interventions, they reported a moderate to large mean ES ($d = 0.72$) at post-treatment and ($d = 0.64$) at follow-up (range: 1 month – 2 years post-treatment, median = 7 weeks). Interestingly, this result was in stark contrast to the mean ES ($d = 0.19$) for the 14 pharmacological trials.

Recent meta-analyses and reviews of meta-analyses of youth psychotherapy with improved methodological rigor (e.g., inclusion of intent-to-treat analyses, randomized controlled trials, and unpublished dissertation research; Klein, Jacobs, & Reinecke, 2007; Weisz, McCarty, & Valeri, 2006; Zirkelback & Reese, 2010) have found mostly small to moderate treatment ESs. For example, Klein et al. (2007) conducted a meta-analysis of 11 randomized controlled trials of CBT for youth meeting diagnostic criteria for unipolar depression. The mean ES for the six studies including an intent-to-treat analysis (i.e., including outcomes of all clients initially randomized into conditions) was 0.26, while the much larger mean ES (0.94) for the remaining five studies only compared treatment group completers to control group participants. Weisz et al.'s (2006) meta-analysis of psychotherapy for youth diagnosed with depression resulted in a mean treatment ES of 0.34 for 35 randomized controlled studies—including peer reviewed studies, non-peer-reviewed studies (e.g., book chapters), and doctoral dissertations.

Despite these overall positive findings for clinical trial studies, psychotherapy effectiveness studies evaluating treatment-as-usual care with youth have been mixed (e.g., Bickman, 1996; Garland et al., 2013; Weisz et al., 2005) calling into question how well evidence-based therapies from randomized clinical trials perform when implemented

in many “real world” contexts. As previously noted, most of the efficacy research has focused on treatment completers which tends to increase effect sizes and may not represent outcomes in real world settings. For example, premature termination rates are staggeringly high in real world settings. Decades of research show that despite the accumulation of hundreds of evidence-based treatments for child and adolescent behavior problems, approximately half of the families with children receiving mental health services continue to terminate treatment prematurely (Gould, Shaffer, & Kaplan, 1985; Harpaz-Rotem, Leslie, & Rosenheck, 2004; Weisz, Weiss, & Langmeyer, 1987; Wierzbicki & Pekarik, 1993). One serious implication of this high premature termination rate is more costly services in the future due to unresolved symptoms (Farmer & Burns, 1997). The challenges faced in psychotherapy with youth in general are even more sobering when looking at psychotherapy with youth in poverty in particular.

Psychotherapy With Youth in Poverty

In 2002, the President’s New Freedom Commission on Mental Health concluded that “America’s mental health service delivery system is in shambles” and that it was “incapable of efficiently delivering...effective treatments” (p. ii). Since then, the mental health service delivery system in the United States for financially disadvantaged youth receiving mental health care and substance abuse care in community-based PBH programs may still be in shambles (Garland et al., 2013). Financially disadvantaged youth have significantly higher premature termination rates (Wierzbicki & Pekarik, 1993), significantly larger treatment outcome deterioration rates (14% and 24%, respectively; Warren et al., 2010), and the effect sizes are often near zero (Farahmand, Grant, Polo, Duffy, & DuBois, 2011; Lipsey & Wilson, 1998; Weiss, Catron, Harris, &

Phung, 1999; Weisz, 2004). Although single effectiveness studies may sometimes demonstrate larger effect sizes in small samples of youth in PBH settings (e.g., Lee, Horvath, & Hunsley, 2013), no convincing evidence exists of effective routine care in PBH settings on a large scale (Garland et al., 2013). This lack of effectiveness evidence represents a significant area needing improvement. Compared to efficacy trials, the youth involved in PBH settings are more likely to be diagnosed with co-occurring disorders, in families reporting lower incomes, in ethnic minority families, non-insured, designated as disabled by Social Security, and served by therapists with full caseloads and less utilization of evidence-based practices and treatment (Brookman-Frazee, Haine, Baker-Ericzén, Zoffness, & Garland, 2010; Ehrenreich-May et al., 2011; Southam-Gerow et al., 2008; Weersing & Weisz, 2002; Weiss et al., 1999; Weiss, Harris, Catron, & Han, 2003). For example, Brookman-Frazee, Haine, Baker-Ericzén, Zoffness, & Garland (2010) found that youth from lower SES homes had not only significantly worse treatment outcomes but also received less quality mental healthcare as evidenced by the utilization of evidence-based practices by their mental healthcare providers. Similarly, in a longitudinal analysis of 62 clinics in California, Zima et al. (2005) found that for youth insured by Medicaid, the median annual income of the clinic county predicted the quality of care (i.e., completeness of clinical assessment, appropriate linkage to other service sectors, patient protection, initiating medication referral, and parental involvement) where clinics in counties below the state median annual income (\$35,725) had significantly lower ratings of quality care than richer counties. The evidence for a system in shambles continues to stack up.

Warren et al. (2010) measured treatment outcomes of youth ($N = 936$) in a PBH setting and found that fewer than half (44%) of youths either improved or recovered, while 32% demonstrated no reliable change and 24% deteriorated. These findings were in contrast to the statistically greater rates of change in a comparison group of youth receiving psychotherapy in a managed care setting. In a similar study, Manteuffel, Stephens, Sondheimer, and Fisher (2008) studied the outcomes of 3,613 youth in 45 different PBH agencies in the United States between 1997 and 2006. The most frequent diagnosis given was a mood disorder (44.4% of youth aged 14 to 15, 38.6% for youths aged 16 to 17, and 33.3% of youths aged 18). The authors used a reliable change index and found, on average that only 36% of youth improved, 50% exhibited no reliable change, and the remaining 14% had deteriorated outcomes.

Another difficulty is the evaluation of mental health treatment for youth in poverty in naturalistic settings. Efficacy trials provide greater control and strengthen internal validity as an experiment, yet lose generalizability in the process. Effectiveness studies, on the other hand, may at times produce large effect sizes, but without control groups (i.e., waitlist, treatment-as-usual, placebo), the internal validity of the studies is weakened, thereby limiting knowledge of whether the change was due to the treatment versus regression to the mean (i.e., clients presenting with highest levels of distress at intake; Lambert & Bickman, 2004). Hence, a research-practice gap persists between efficacy trials and effectiveness studies.

Benchmarking studies for youth in poverty. Benchmarking helps bridge the research-practice gap between controlled clinical trial data and effectiveness studies in clinical settings where almost all psychotherapy is done (Minami, Serlin, et al., 2008).

Benchmarking originally developed as a strategy to evaluate business practices (e.g., Camp, 1989) and is now commonly used in a range of organizations to evaluate the quality of services. Benchmarking is a method of direct statistical comparison between psychotherapy outcomes (i.e., to what degree treatment helped an actual client) in real world clinical settings and those from treatments found to follow best practice standards (i.e., efficacy through testing in clinical trials). The American Psychological Association's (2013) resolution on psychotherapy effectiveness references benchmarking as a viable strategy to compare routine care-based and clinical trial-based psychotherapy outcomes. To date, three benchmarking studies have examined the psychotherapy outcomes of youth from mostly economically impoverished backgrounds. Findings from these studies have been mixed. Although they include small samples of youth involved in community mental health settings in different countries, limiting generalizability, the benchmarking process is exemplified, and two of the three studies show positive results compared to the majority of treatment outcome research for this population. Each of these studies will be reviewed next.

Weersing and Weisz (2002). Weersing and Weisz (2002) conducted a benchmarking study comparing the treatment outcomes of ethnically diverse youth diagnosed with depression at six community mental health centers in the Los Angeles area with outcomes derived from a meta-analysis of clinical trials. Benchmarking methodology allowed Weersing and Weisz to statistically compare treatment outcomes between their study and efficacy studies (i.e., clinical trials). For their meta-analysis, they searched databases (i.e., PsycINFO, PsychLit, and Medline), book chapters, reference lists of articles, and meta-analyses for all published randomized clinical trials in

the English-language on the treatment of youth diagnosed with depression. They limited their meta-analysis to clinical trials of CBT for youth diagnosed with depression giving the rationale that CBT programs were the only treatments with enough efficacy support—resulting in a total of 15 CBT conditions found in 13 clinical trials. Next, they aggregated the treatment outcome results creating benchmarks at the following assessment times: intake (CBT: $K = 15$; Control: $K = 11$), post-treatment (CBT: $K = 15$; Control: $K = 11$), 1- to 3-month follow-up (CBT: $K = 9$; Control: $K = 4$), 4- to 6-month follow-up (CBT: $K = 5$), 7- to 9-month follow-up (CBT: $K = 3$), and 10- to 12-month follow-up (CBT: $K = 9$). They converted depression scores into normative z -scores for both their sample and the clinical trial data since various measures of depression were utilized. Since the youth in their sample had varying treatment lengths, Weersing and Weisz calculated the mean of the z scores for all available assessments within the same range of the benchmark period: intake ($n = 67$), 6-month ($n = 37$), and 1-year ($n = 35$). Then, they evaluated whether the z -score mean from their sample fell within the 95% confidence interval for either the control benchmark mean or the CBT benchmark mean. In order to estimate a 1- to 3-month follow-up mean symptom trajectory for their sample, they used hierarchical linear modeling.

Their findings are notable. The mean z -scores they found at 3-month follow-up ($M = 1.23$) indicated that the mean symptom severity of youth in their sample was almost identical to that of the efficacy benchmark no-treatment control group ($M = 1.24$), and contrasting markedly from the CBT group symptom severity ($M = -0.13$; lower numbers indicating less symptom severity). In other words, they found that the youth at the community mental health center had treatment outcomes similar to symptom outcome

trajectories found for no-treatment control groups in clinical trials up until 3-months post-treatment. Additionally, the 6-month follow-up comparison showed that the mean z-score for their sample ($M = 0.84$) was not found to be within the confidence intervals of the CBT treatment benchmark range (95% CI [0.28 to -0.46]). Finally, at 1-year follow-up assessment, the community mental health center sample did report a z-score mean ($M = 0.18$) within the confidence interval of the CBT treatment group benchmark (95% CI [0.18 to 0.50]), but the small sample ($N = 35$) and limited corresponding benchmark study sample inclusion ($N = 1$) limit the generalizability of this last finding.

Curtis, Ronan, Heiblum, and Crellin (2009). Curtis, Ronan, Heiblum, and Crellin (2009) evaluated the transportability of Multisystemic Therapy (MST) for the treatment of juvenile offenders ($N = 65$) in a community-based clinic in New Zealand and found evidence of effectiveness. The most frequent reasons for referral included verbal/physical aggression at home, school, or in the community (60%, $n = 39$), truancy (14%, $n = 9$), substance abuse (8%, $n = 5$). They conducted a one-group pre-post test design. In order to compare their single sample to best practice benchmarks, they chose a meta-analysis (Curtis et al., 2004) for completion rate (benchmark of 86%) and three RCTs (Borduin et al., 1995; Henggeler, Melton, Brondino, Scherer, & Hanley, 1997; Henggeler, Melton, & Smith 1992) for the other outcome measures they labeled “ultimate outcome data” (i.e., frequency and severity of truancy and offending behaviors) and “instrumental outcome data” (i.e., client overall adjustment and behavioral change, youth compliance, family communication, and family relations).

For the three RCTs utilized in creating effect size benchmarks, instrumental outcome effect size estimates were $d = 0.36$, $d = 0.14$, and $d = 0.04$ respectively. Curtis

et al. found the following pre-post treatment effect sizes for their sample: $d = 0.32$ for ultimate outcomes and $d = 0.75$ for instrumental outcomes, aggregating the overall effect size for treatment at $d = 0.53$. MST treatment on average lasted 155 days ($SD = 39.22$) with the range of 61 to 253 days and a high completion rate (98%). Through their study, they were able to show the successful transportability of an evidence-based treatment into a publically-funded community mental health setting.

Yet, some methodological aspects of their study make it less comparable to effectiveness studies of most routine mental health care for youth from economically impoverished backgrounds. First, we don't know the specific SES of each youth in the study. The only SES factor reported was partial and not at the individual level: sixty-nine percent of clients ($n = 45$) lived in the most deprived areas of New Zealand (mean household income = \$17,700). Secondly, the benchmark creation process is questionable. The three RCTs selected did not result from a systematic search, and the only reported criterion to create benchmarks was that the MST clinical trials included chronic juvenile offenders. No systematic search for more recent clinical trials or unpublished dissertations was conducted. This point is considerable given the overall small effect sizes reported by the RCTs utilized as a "best practice" benchmark.

Lee, Horvath, and Hunsley (2013). Finally, Lee, Horvath, and Hunsley (2013) compared seven effectiveness studies evaluating evidence-based treatments (i.e., Interpersonal Process Psychotherapy and Cognitive Behavioral Therapy) for youth from around the world against benchmarks derived from published meta-analyses. The outcomes being compared were completion and improvement rates for what they categorized as internalizing disorders (i.e., depression, mixed anxiety disorders, and

Obsessive-Compulsive Disorder). Five of the seven effectiveness studies were from youth at community mental health centers (CMHCs) in five different countries and are salient for this discussion.

In comparison to the efficacy benchmarks, the CMHC studies had mixed results, but the majority of results were equivalent to the benchmarks. The rates of completion and improvement for two CMHC studies on Obsessive Compulsive Disorder (Farrell et al., 2010; Valderhaug et al., 2007) were statistically equivalent to the efficacy benchmark. A depression study comparison was even more impressive. The depression efficacy benchmark utilized (Watanabe et al., 2007) had a completion rate of 85.8% and an improvement rate of 49.6%. Surprisingly, the completion and improvement rates (100% and 70%, respectively) of the CMHC depression study in the United States (Weisz et al., 2009) were significantly greater.

The benchmark studies addressing mixed anxiety disorders yielded inconsistent results. The mixed anxiety efficacy benchmark (In-Albon & Schneider, 2007) consisted of completion rates (individual therapy, 84.9%; family therapy, 82.9%) and client improvement rates (individual therapy, 72.1% & family therapy, 76.9%). Lau et al. (2010) found rates of completion and improvement for their CMHC sample on mixed anxiety to be statistically equivalent to these benchmarks. Bodden et al.'s (2008) CMHC study for youth diagnosed with mixed anxiety disorders had a significantly higher completion rate (96.8%) than the mixed anxiety benchmark for individual therapy, but significantly lower improvement rates (42.0%) for family therapy.

Limitations. Two limitations in all three benchmark study findings are noteworthy. First, all the aforementioned benchmark studies included small effectiveness

study samples, ranging from 28 to 67 participantsvb. Secondly, the authors did not evaluate the equivalency in outcome measures between benchmarks and effectiveness studies—the person (e.g., client, parent, or therapist) completing the measure and the specificity of the measure have both been shown to produce greater or lesser effect size estimates (Minami, Wampold, Serlin, Kircher, & Brown, 2007).

Summary of problem one. To summarize, youth from economically impoverished backgrounds typically incur a disproportionate number of life stressors, yet they often receive inadequate services (Brookman-Frazee et al., 2010; Zim et al., 2005) and are most at-risk for ongoing mental health problems that are costly on an individual, familial, and community level (Frank & Glied, 2006). Even though several meta-analyses of randomized clinical trials for this population have found treatment outcome effect sizes ranging from 0.34 (Weisz et al., 2006) to 1.27 (Lewinsohn & Clarke, 1999), the effectiveness studies of routine care with youth in community-based publically funded PBH settings are mixed. Although some recent benchmarking studies have shown the successful transportability of evidence-based treatments to settings with small samples of financially disadvantaged youth (e.g., Lee et al., 2013), many more effectiveness studies have shown psychotherapy to have minimal clinical impact (e.g., $d = 0.25$, Farahmand et al., 2012; $d = 0.08$, Farahmand et al., 2011; $d = 0.12$, Lipsey & Wilson, 1998; Weersing & Weisz, 2002; $d = 0.08$, Weiss, Catron, Harris, & Phung, 1999).

Measuring Progress and Alliance for Quality Improvement

Internationally, policymakers and clinical scientists are calling for the use of outcome measures in routine mental healthcare (Bohanske & Franczak, 2010; Lambert,

2010; Wolpert, Cheng, & Deighton, 2014). Perhaps most influential is the National Quality Strategy of the 2010 Patient Protection and Affordable Care Act, which called for quality outcome measures as a strategy to improve health care services (Zima & Mangione-Smith, 2011). Outcome measures may be completed by the client, peer, parent, care-giver, or clinician. They may be specific to a diagnosis or problem area or they may be broad-based assessments of global distress or wellbeing. Today, more clinical scientists are seeing the central importance of self-report measures—systematically soliciting a young person’s own perspective (Bickman, 2008; Deighton et al., 2014; Duncan, Sparks, Miller, Bohanske, & Claud, 2006). Outcome psychotherapy measures used in this way not only measure treatment outcomes, but can also become clinical tools to routinely inform therapists about session-to-session progress and other dimensions of treatment (e.g., therapeutic alliance).

Client feedback. Outcome measures that frequently solicit feedback directly from the clients to inform the course of treatment may be referred to as progress monitoring, outcome monitoring, client feedback, measurement-feedback systems, stepped care, practice-based evidence, or client-directed outcome-informed among others (e.g., Bickman, 2008; Duncan, Miller & Sparks, 2004; Goodman, McKay, & DePhilippis, 2013; Lambert & Shimokawa, 2011; Overington & Ionita, 2012). Here, I will refer to these collectively as client feedback due to the term indicating the client’s perspective, and it being a broadly used term—especially when discussing those client feedback systems with the most evidence of effectiveness (Lambert & Shimokawa, 2011).

Client feedback has a well-established and steadily growing evidence base in improving treatment outcomes of adults when compared to treatment-as-usual (Anker et

al., 2009; Reese, Norsworthy, & Rowlands, 2009; Reese et al., 2010; Schuman et al., 2015; Shimokawa, Lambert, & Smart, 2010; Slone et al., 2015. For example, Shimokawa et al. (2010) conducted a meta/mega-analysis of six clinical trials of the Outcome Questionnaire-45 (OQ-45; Lambert et al., 1996) client feedback system with over 4,000 individual therapy clients. They found in an intent-to-treat analysis that, for clients who were identified as at-risk for treatment failure, the use of client feedback (compared to treatment-as-usual) resulted in the following aggregated between-group effect sizes and (rates of reliable/clinically significant improvement): $g = -0.28$ (30.9%) when feedback was only seen by therapist, $g = -0.36$ (38.7%) when feedback was seen by both therapist and client, and $g = -0.44$ (37.6%) when feedback to the therapist was accompanied by clinical support tools which make suggestions for resolution of identified problems. Effect sizes were weighted (Hedges's g ; random effects model) and negative effect sizes indicated lower distress levels. They also found that the client feedback improved outcomes between clients on-track for treatment success. Feedback provided to the therapist resulted in a between-group effect size of $g = -0.12$ and an odds ratio of $OR = 1.20$ for the occurrence of reliable/clinically significant improvement at termination. Better yet, when feedback was seen by both client and therapist, between-group effect size was $g = -0.18$ and an odds ratio was $OR = 1.65$ for the occurrence of reliable/clinically significant improvement at termination.

Results from other adult studies comparing client feedback to treatment-as-usual groups are similarly positive. Several adult studies have shown consistently positive results supporting PCOMS. In a recent meta-analysis of PCOMS, Lambert and Shimokawa (2011) evaluated the outcomes of 558 adults and reported that those in the

client feedback group had less than half the odds of experiencing deterioration and 3.5 times higher odds of experiencing reliable change. In a RCT of couples therapy, Reese et al. (2010) found that four times as many couples in the feedback condition reported clinically significant change at the end of treatment compared to couples receiving treatment-as-usual. Couples in a feedback condition also reported improved psychological well-being more rapidly. Two group psychotherapy studies of PCOMS also reported higher rates of reliable and clinically significant change, more group session attendance, and significantly larger pre-post treatment therapy gains when compared to treatment-as-usual conditions ($d = 0.28$, Schuman et al., 2015; $d = 0.41$, Slone et al., 2015, respectively).

Client feedback studies with youth. The research evidence is compelling with adults, yet few studies have evaluated the benefit of client feedback with youth. The extant research, though limited, shows promising results. In one study, youth provided with more frequent opportunity to give feedback to their clinicians were shown to have faster rates of change (Nelson, Warren, Gleave, & Burlingame, 2013). In another study, Bickman, Kelley, Breda, de Andrade, & Riemer (2011) conducted a randomized clinical trial of youth psychotherapy at 25 sites of a private, for-profit, behavioral health organization. Client feedback in this study meant providing clinicians mean scores on the Symptoms and Functioning Severity Scale (SFSS; Bickman et al., 2010) as assessed by the youth, care-giver, and clinician. They also received alerts if client symptoms ranked within the 25th percentile of severity. Youth in the treatment-as-usual control group received services where clinicians received 90-day cumulative feedback. In the experimental group, clinicians received weekly reports of client feedback in addition to

90-day cumulative client feedback and saw a feedback effect size of 0.17 over the control group. They reported, though, that less than half (46%) of these reports were available to clinicians within a week of each session and client feedback reports were available to clinicians a median of nine days ($M = 12.3, SD = 22.3$) after the end of each session. However, when Bickman and colleagues conducted a dose-response analysis of client feedback reports viewed by clinicians, treatment effect sizes increased by an additional 50%. The results indicated that agency level implementation of client feedback significantly improved outcomes without the costly or lengthy implementation of empirically supported treatments as others have pointed out (e.g., Laska, Gurman, & Wampold, 2013; McHugh & Barlow, 2012).

A recent community-based effectiveness study (Cooper, Stewart, Sparks, & Bunting, 2013) showed even more promising results where researchers evaluated the use of a client feedback system called the Partners for Change Outcome Management System (PCOMS; Duncan, 2012) in a public school-based counseling context with youth in Ireland. Findings yielded an overall effect size of $d = 1.49$ (Cooper et al., 2013). Although nascent for youth psychotherapy, the burgeoning client feedback research activity indicates that client feedback systems are empirically supported both in efficacy and effectiveness studies—with the greatest improvement occurring when clinicians view the client’s progress every session.

Client feedback and evidence-based practice. Two client feedback systems (Outcome Questionnaire 45.2, OQ-45.2; Lambert et al., 1996; and PCOMS; Duncan, 2012) are recognized as evidence-based interventions by the Substance Abuse Mental Health Services Administration (SAMHSA) and listed in the National Registry of

Evidence-based Programs and Practices (NREPP). Besides being an evidence-based practice itself, client feedback stands out in being a flexible way to facilitate the adaptation of other empirically supported treatments in a real world, clinical setting (Garland et al., 2014). Therefore, client feedback becomes an ecologically valid and promising quality improvement strategy in PBH settings where money and resources are limited. At the same time, client feedback is instrumental in carrying out what the Presidential Task Force on Evidence-Based Practice of the American Psychological Association (APA) considers evidence-based practice: "[T]he integration of the best available research with clinical expertise in the context of patient [sic] characteristics, culture, and preferences" (APA Presidential Task Force on Evidence-Based Practice, 2006, p. 273). When clients have the continual opportunity to voice their values and express their preferences, then clinicians can more accurately adjust their treatment plan contextually for each unique client.

Client feedback and the working alliance. The working alliance refers broadly to the affective and collaborative aspects of the relationship between the therapist and the client. The concept of an alliance was originally conceptualized in the psychoanalytic field as the client's trusting and affectionate feelings toward the therapist (Wampold, 2001). Bordin (1979) expanded the concept outside of psychoanalysis, calling it the *working alliance*. Bordin's more collaborative conceptualization of the alliance contains three main components—*goals*, *task*, and *bond*—where agreement on the therapeutic tasks and goals along with strong relational bond are essential to a working alliance. This transtheoretical conceptualization of alliance has been widely used in psychotherapy research (Horvath, Del Re, Flückiger, & Symonds, 2011) and is the most frequently cited

common factor (Wampold, 2001). When measured and included as a variable, the working alliance is often found to be a robust predictor of psychotherapy outcomes (Horvath et al., 2011; Horvath & Symonds, 1991; Martin, Garske, & Davis, 2000; Orlinsky, Rønnestad, & Willutzki, 2004)—accounting for more variance in psychotherapy outcomes than the specific therapeutic approach (Wampold, 2001).

Specific to therapy with youth, the working alliance is also supported by evidence that the three components of the alliance (goals, task, and bond) between clinicians, children and their families significantly contributes to clinical outcomes (Karver, Handelsman, Fields, & Bickman, 2006; Lambert, 2007; McLeod, 2011; Shirk & Karver, 2003). Alliance has also been shown to be significantly related to youth psychotherapy for attendance (Connors, Carroll, DiClemente, Longabaugh, & Donovan, 1997; Garcia & Weisz, 2002). Mental health professionals working with youth have responded positively to brief measures being used to measure alliance (Law & Wolpert, 2014; Miller, Duncan, Brown, Sorrell, & Chalk, 2006; Timimi, Tetley, Burgoine, & Walker, 2013). Similarly, qualitative studies have shown that youth place much value on the therapeutic alliance (Day, Carey, & Surgenor, 2006; Stith, Rosen, McCollum, Coleman, & Herman, 1996).

As an established evidence-based practice, PCOMS is unique in that it not only routinely monitors the clients' level of distress in session, but it also routinely evaluates the working alliance. By including the working alliance as a major component, PCOMS fulfills the recommendation from the APA's Division 29 Task Force on Empirically Supported Relationships that clinicians monitor outcome and the therapeutic alliance on an ongoing basis (Ackerman et al., 2001). Next, I will describe PCOMS in more detail as it relates to the current study.

The Partners for Change Outcome Management System (PCOMS). To meet the urgent need in improving mental health services for youth from economically impoverished backgrounds, the Partners for Change Outcome Management System (PCOMS; Duncan, 2012) may be a promising option. Both the Campaign for Mental Health Reform (CMHR; 2005) representing leadership of 16 national health organizations and the President's New Freedom Commission on Mental Health (2002) proposed comprehensive reform agendas that pointed to client feedback principles including (a) consumer-centered services, (b) common factors in care, and (c) recovery in daily functioning instead of cure of illness (as summarized in Bohanske & Franczak, 2010). PCOMS is designed with each of these three principles in mind and the empirical research so far is encouraging.

For example, PCOMS was recently evaluated (Reese, Duncan, Bohanske, Owen, & Minami, 2014) as a quality improvement strategy for psychotherapy with adults at or below 100% of the federal poverty level at a large PBH agency in Southwestern United States. They observed a treatment effect size of $d = 1.34$ for adults in treatment for depression treatment ($n = 1,589$). Employing a benchmarking methodology, (Minami, Serlin, et al., 2008) they reported that their sample effect size easily surpassed the clinical trial benchmark effect sizes ($d = 0.84$ for completer samples and $d = 0.93$ for intent-to-treat samples) for adults diagnosed with depression.

PCOMS consists of two adult measures and alternative child measures. The 4-item Outcome Rating Scale (ORS; Miller et al., 2003; see Appendix A) consists of four 10 cm visual analogue scales, assesses three areas of subjective distress (e.g., individual, interpersonal, and social), and is completed by the client at the beginning of each

session. The 4-item Session Rating Scale (SRS; Duncan et al., 2003) consists of four 10 cm visual analogue scales, assesses the working alliance (e.g., relationship, goals and topics, approach or method, overall), and is completed by the client at the conclusion of each therapy session. The Child Outcome Rating Scale (CORS) and Child Session Rating Scale (CSRS; Duncan et al., 2006) were developed for youth (aged 6-12) to function as alternatives to the ORS and SRS for older youth and adults (13 years and older; Duncan et al., 2003). The four items of the CORS and CSRS measure similar domains as the ORS and SRS, but substitute item descriptions with more age appropriate language and use smiley or frown faces as anchors (see Appendix B). The use of these emoticons adds to the suitability of these scales for youth of various ethnic origins.

Scores from all four measures have been tested and shown adequate psychometric properties (e.g., reliability and validity) considering their brief nature designed for routine use in clinical settings. The ORS was developed as a brief alternative to the lengthier but well-validated OQ-45 (Lambert et al., 1996). Psychometric testing has shown the ORS and CORS to have adequate reliability and validity (Bringham, Watson, Miller, & Duncan, 2006; Campbell & Hemsley, 2009; Duncan et al., 2006; Miller, Duncan, Brown, Sparks, & Claud, 2003). Similarly, the SRS has been repeatedly tested for reliability and construct validity and shows adequate measurement of a single global alliance construct (Anker, Owen, Duncan, & Sparks, 2010; Campbell & Hemsley, 2009; Duncan et al., 2003; Reese et al., 2010). These psychometric studies will be elaborated further in the methods section.

The PCOMS has also proven itself to have a high degree of feasibility as a clinical intervention. First, it takes an a-theoretical (i.e., common factors) stance in

measuring psychotherapy outcome and alliance. Second, each measure takes two minutes or less to administer, score, and discuss (Duncan, 2012). Third, measures can be implemented in paper, oral, or electronic form; paper measures are freely available for download for individual use at www.heartandsoulofchange.com or www.whatispcoms.com. Fourth, PCOMS has face validity in being more widely accepted by therapists than longer measures (Duncan et al., 2003). For example, when the 4-item SRS was compared to the 12-item Working Alliance Inventory-Short (WAI-S; Tracey & Kokotovic, 1990) between two clinics, the SRS had a utilization rate of 96% compared to WAI-S's rate of 29% (Duncan et al., 2003). Similarly, when the 4-item ORS was compared to the 45-item OQ-45 (Lambert et al., 1996) at two similar clinical sites, the ORS achieved an 86% compliance rate whereas the therapist compliance rate of the OQ-45 was only 25% (Miller et al., 2003).

Finally, PCOMS is being implemented internationally in community-based and PBH settings as an evidence-based practice (e.g., Bluegrass Regional Mental Health and Southwest Behavioral Health Services in the United States, Saskatoon Health Region in Canada, Bufetat in Norway, Lincoln-shire Child and Adolescent Mental Health Services in the United Kingdom, and Wesley Community Action in New Zealand), and as of 2012, over 100,000 clients were annually participating with the PCOMS (Duncan, 2012).

Summary of problem two. Despite burgeoning evidence of PCOMS being effective as a quality improvement strategy in the adult psychotherapy literature, and more specifically, in PBH (Reese et al., 2014), systematic evaluation of PCOMS with youth in community-based PBH settings is lacking. PCOMS has neither been

systematically evaluated in a naturalistic setting with youth in the United States, nor youth in the United States from economically impoverished backgrounds.

Purpose of This Study

The current study was designed to answer two questions: (a) in comparison to efficacy trial benchmarks, is psychotherapy utilizing continuous client feedback (e.g., PCOMS) effective in reducing depression-related distress among youth in poverty within a PBH setting?; and (b) in comparison to efficacy trial benchmarks, is psychotherapy utilizing continuous client feedback (e.g., PCOMS) effective in reducing overall psychological distress among youth in poverty within a PBH setting? These questions come from the well-documented findings in the literature that youth from economically impoverished backgrounds are more susceptible to have mental health problems (e.g., Frank & Glied, 2006) and worse outcomes of mental health services than youth who do not face economic impoverishment (e.g., Manteuffel et al., 2008; Warren et al., 2010; Wierzbicki & Pekarik, 1993).

Benchmarking methodology (Minami, Serlin, et al., 2008) will be utilized to address these questions. First, following previous benchmarking studies (e.g., Minami et al., 2007), efficacy benchmarks will be constructed from results found in clinical trials of bona fide treatment groups, control groups, and treatment-as-usual groups. Second, the effectiveness of psychotherapy provided to youth at a PBH agency will be evaluated by comparing the observed pre-post effect size estimate against these efficacy benchmarks. I hypothesize that the psychotherapy outcomes at the PBH sample will be (a) clinically equivalent to outcomes of the treatment efficacy condition observed in the clinical trials and (b) superior to wait list controls and treatment-as-usual comparison groups.

Chapter Two: Method

The current study is an evaluation of the effectiveness of psychotherapy in a PBH agency using client feedback provided to youth from economically impoverished backgrounds. In this chapter, I first describe the current PBH sample, procedures and treatment outcome measures. Second, I describe benchmarking methodology and construct the benchmarks. Third, I detail the data analyses including effect size calculations and benchmarking procedures. Finally, I specify each hypothesis.

Design

The current naturalistic study utilized a benchmarking design to evaluate the effectiveness of a PBH agency that has employed client feedback with youth clients. Benchmarking allows the comparison of treatments delivered in naturalistic non-controlled settings against reliably determined effect size estimates (ESs) in single clinical trials or meta-analyses of clinical trials (Minami et al., 2007; Weersing & Weisz, 2002). Effect size estimates (ESs) were calculated for this sample of youth and then compared to constructed benchmarks (i.e., no treatment/waitlist control group and intent-to-treat group) derived from treatment outcome measures used in efficacy trials with youth. Intent-to-treat samples from efficacy trials include scores from clients who terminate prematurely and therefore are more similar to real world clinical samples than samples only reporting effect size estimates based on completer samples. Also, since naturalistic clinical settings often lack a no-treatment group as a control, using a waitlist (i.e., no-treatment) control group ES benchmark will allow assessment for effectiveness testing with a comparison group and therefore strengthen internal validity.

Procedures

Treatment outcome data came from the psychotherapy archives of Southwest Behavioral Health Services (SBHS), a not-for-profit, comprehensive PBH organization serving a diverse range of individuals and families in Maricopa (Phoenix), Mohave, Yavapa, Coconino, and Gila counties in Arizona. The Institutional Review Board (IRB) at the University of Kentucky determined this project to be exempt from IRB review (see Appendix C). SBHS annually serves roughly 6,800 youth and 16,600 adults in both urban and rural settings, and clinicians use PCOMS comprehensively throughout its locations. They provide clinical services, including mental health and substance abuse treatments, to Medicaid-insured youth and adults at or below 100% of the federal poverty level. SBHS included PCOMS in the treatment with all youth clients involved in this study. Youth 13 to 17 years old completed the ORS and youth 6 to 12 years old completed the CORS. SBHS therapists were trained in PCOMS processes over two days (12 hours) and then received annual one-day booster trainings. Agency-wide quality improvement policies required that therapists collect outcome data and routinely identify and discuss at-risk clients in ongoing supervisory meetings.

SBHS required clinicians to use PCOMS, but they did not mandate or monitor their specific treatment approaches. Therapists ($N = 86$) were predominantly female (84.2%) and were African American (2.1%), Latino(a)/Hispanic (9.8%), and Caucasian (88.1%). Roughly two-thirds of therapists (68.2%) had degrees in the counseling field, and the remaining third had degrees in clinical social work (12.7%), substance abuse

counseling (11.3%), and psychology (9.4%). Therapists were all licensed and had a master's degree or higher.

Participants

For the current study, SBHS granted permission for data analysis from youth discharged cases between January 2008 and March 2014. The initial dataset included 4,558 cases, to which three sets of deletions were made: (a) one duplicate case, (b) 126 cases where the client was older than 17 years old at intake, and (c) 42 cases where the client was younger than 6 years old at intake. The remaining clients ($N = 4,389$) were predominantly White/Euro-American (32.3%), male (50.3%), and ranged in age from 6 to 17 ($M = 12.12$, $SD = 3.28$). As reported in Table 1, the full sample also included Latino(a)/Hispanic (22.8%), Black/African American (7.0%), Native American (1.3%), and other ethnicities (2.3%). Sociodemographic information for youth (e.g., age, sex, and ethnicity) from the depression sample is also presented in Table 1.

Table 1

Client Demographic Information for Full and Depression Samples

	Full sample (<i>N</i> = 4,389)	Depression-related sample (<i>N</i> = 469)
Age <i>M</i> (<i>SD</i> , Range)	12.12 (3.28, 11)	12.88 (2.97, 11)
Female <i>n</i> (%)	2,165 (48.6)	318 (67.8)
Male <i>n</i> (%)	2,207 (50.3)	151 (32.2)
Sex Unknown <i>n</i> (%)	17 (0.4)	0 (0)
Latino(a)/Hispanic <i>n</i> (%)	1,002 (22.8)	109 (23.2)
African American <i>n</i> (%)	309 (7.0)	19 (4.1)
Native American <i>n</i> (%)	55 (1.3)	9 (1.9)
Euro-American <i>n</i> (%)	1,416 (32.3)	138 (29.4)
Other Ethnicity <i>n</i> (%)	102 (2.3)	11 (2.3)
Unknown Ethnicity <i>n</i> (%)	1,505 (34.3)	183 (39.0)

Following the inclusion criteria of Reese et al. (2014), I included youth in the depression sample who (a) had intake ORS/CORS scores below the clinical cut-off (i.e., < 28 for adolescents 13-17 years of age and < 32 for children 6-12 years of age) representative of a clinical population, and (b) completed at least two psychotherapy

sessions (clients with both a pre and post treatment score) in order to calculate a score difference. These inclusion criteria were also consistent with the efficacy trial samples from Weisz et al. (2006). Primary diagnoses were determined by therapists by the third session. Included in the PBH depression sample were youth diagnosed with major depressive disorder, dysthymic disorder, depressed mood NOS, adjustment disorder with depressed mood, and adjustment disorder with mixed anxiety and depressed mood. Information about medication usage and comorbidity was unavailable.

Measures

The two outcome measures (ORS and CORS) utilized in evaluating the effectiveness of psychotherapy with the current sample are from PCOMS. All clients included in the current study participated in PCOMS treatment.

Outcome Rating Scale (ORS). The ORS (see Appendix A; Miller et al., 2003) is an ultra-brief 4-item self-report outcome measure included in PCOMS (Duncan, 2012). The four items are visual analogue scales (VAsSs) and measure individual, social, interpersonal, and overall psychological distress--areas widely considered to be indicators to track successful treatment outcome (Hill & Lambert, 2004). Scores from each of the four visual analogue scales are summed to give a summative value from 0 to 40 cm.

The ORS can evaluate treatment outcomes based on reliable change or clinically significant change. The ORS has a clinical cutoff score of 25, a value established from a sample ($N = 34,790$) consisting of clients of low SES from a community mental health center (Miller et al., 2003) and an alcohol and drug treatment center (Miller, Mee-Lee, Loum, & Hubble, 2005). Likewise, the ORS has a clinical cutoff score of 28 for clients

aged 13-17 (Duncan et al., 2006). A score less than 25 (or 28 for adolescents) indicates scores typical for clinical populations.

Using the Reliable Change Index (RCI; Jacobson & Truax, 1991), Duncan (2014) determined reliable change on the ORS to be a change of 6 points. Combining the RCI criteria and clinical cutoff score, the ORS can say that a client initially scoring below 25 (28 for adolescents) will experience reliable and clinically significant change by scoring 6 points higher and scoring at or above the respective clinical cutoff score.

Available in both electronic and paper-based formats, the ORS is used as a clinical tool in the presence of the therapist. At the beginning of every therapy session, clients rate their levels of distress-wellbeing. Clients make a mark on each of the four VASs that are 10 cm in length, with marks near the left end of the scale indicating low levels of well-being and marks near the right end of the scale indicating high levels of well-being. Lower scores are assumed to indicate less well-being (more distress), while higher scores are assumed to indicate more well-being (less distress).

Independent and dependent variables need to be measured through psychometrically sound assessment instruments, meaning that measures need to have been judged adequate to yield scores that evidence internal consistency and construct validity. In addition to the PCOMS manual (Duncan, 2011), four psychometric studies have evaluated the reliability and validity of the ORS (Bringinghurst et al., 2006; Campbell & Hemsley, 2009; Duncan et al., 2006; Miller et al., 2003). The average Cronbach's alpha (α) coefficients for ORS scores across all four studies was .85 (clinical samples) and .95 (nonclinical samples; Gillaspie & Murphy, 2011). Notably, Duncan et al (2006) reported an average Cronbach's alpha (α) coefficients for ORS scores for youth aged 13-

17 as .93. Three studies (Bringhurst et al., 2006; Campbell & Hemsley, 2009; Miller et al., 2003) evaluated the concurrent validity of ORS scores by comparing ORS scores to the Outcome Questionnaire 45.2 (Lambert et al., 1996), a longer and more established outcome assessment, resulting in an average bivariate correlation of .62 (range .53 to .74; Gillaspay & Murphy, 2011). This moderately strong correlation provides concurrent evidence of validity that scores from the ORS can be seen as an ultra-brief alternative to the longer (45-item) OQ-45.2. Regarding convergent validity, Campbell and Hemsley (2009) found moderate to strong correlations between the ORS total scores and the Rosenberg Self-Esteem Scale ($r = .66$; Rosenberg, 1989), Quality of Life Scale ($r = .74$; Burckhardt & Anderson, 2003), and the three sub-scales of the Depression Anxiety Stress Scale ($r = .46 - .71$; Lovibond & Lovibond, 1995).

Finally, feasibility of using a measure in “real world” clinics is an important element when considering validity of a measure that is both an assessment and a clinical tool. Brown and colleagues (1999) found that only if a measure or combination of measures took less than five minutes to complete, score, interpret, and share/discuss with a client, clinicians were more likely to see the practicality of it. The development of the Outcome Rating Scale (ORS) was in response to a need for a more feasible (i.e., briefer) means for obtaining client feedback and evaluating treatment outcomes (Duncan, 2012). To evaluate the feasibility of the ORS directly, Miller et al (2003) compared therapist compliance rates for utilization of the ORS or OQ-45.2 between two sites with similar clients and similar mandates. After 12 months, they found that the ORS achieved a compliance rate of 89% while the OQ-45.2 maintained a compliance rate of 25%.

Child Outcome Rating Scale (CORS). Like the ORS, the CORS (Appendix B; Duncan et al., 2003) is an ultra-brief 4-item visual VAS self-report outcome measure for progress monitoring and as a clinical tool in session with the mental healthcare provider. The CORS also has gone through psychometric evaluation. Duncan et al. (2006) carried out a 4-year validation study including 20,000 administrations from over 3000 youth—children aged 6-12 using the CORS and adolescents aged 13-17 using the ORS. Coefficient alpha for the CORS was estimated at .84 displaying strong evidence of reliability. In terms of concurrent validity, a Pearson product moment correlation analysis of .61 showed that scores from the CORS significantly correlated with the well-established Youth Outcome Questionnaire (YOQ; Burlingame et al., 2001). Construct validity was also evidenced through findings showing that, while a clinical sample of CORS scores significant increased, the scores from a non-clinical sample only minimally varied from a pre- and post-test (Duncan et al., 2006). The CORS also has a clinical cutoff score of 32 indicating that a score less than 32 is typical for clinical populations for youth 6 to 12 years old (Duncan, 2014).

Benchmark Methodology

Practice-based observational research typically does not allow for comparing treatment groups with a no-treatment control group; thus weakening internal validity. Instead, several researchers have started developing benchmarking techniques in order to compare routinely monitored outcomes over time with established normative samples and meta-analyses of clinical trials. I drew upon benchmarking methodology, which is increasingly being utilized in psychotherapy effectiveness studies (see Lee, Horvath, & Hunsley, 2013; Minami et al., 2009; Minami et al., 2007; Minami, Wampold,

et al., 2008; Reese et al., 2014). As outlined in Minami, Serlin, et al. (2008), the benchmarking strategy requires three steps: (a) construct pre-post benchmarks (i.e., ESs) from randomized controlled trials (RCTs) with waitlist-control and intent-to-treat samples, (b) estimate the pre-post ES of the current sample being evaluated, and (c) statistically compare the current sample ES (i.e., youth in PBH) against the constructed benchmarks derived from RCTs.

Serlin and Lapsley (1985, 1993) proposed a “good-enough principle” allowing for statistical testing with a range-null hypothesis to prevent rejection of a point-null hypothesis due to a large N (i.e., Type I error). Consistent with recent benchmarking studies analyzing large naturalistic data sets (Minami et al., 2009; Minami, Wampold, et al., 2008, Reese et al., 2014), an *a priori* margin of difference of 10% was utilized, indicating a clinically trivial treatment effect (i.e., 90-110% of efficacy trial benchmark ESs). For example, if the treatment group benchmark ES is $d = 1.00$, the range would be 0.9–1.10, indicating a good-enough method of testing clinically meaningful differences with large samples.

A range-null hypothesis (e.g., $H_0 : \delta_{PBHdep} \leq \delta_{ITT} - 10\%$) is used instead of a traditional point null hypothesis (e.g., $H_0 : \delta_{PBHdep} = \delta_{ITT}$). Range-null hypotheses follow a noncentral t statistic (Serlin & Lapsley, 1985; 1993) and a normal distribution is approximated. In order to statistically compare the current PBH sample ESs against the clinical trial benchmark ESs, critical values are calculated to allow for statistical testing with a range-null hypothesis. In other words, critical values are based on this range surrounding the benchmark ESs. For example, a critical value is calculated for the treatment group benchmark ES at $d_{ITT} - 10\%$, where $d_{ITT} - 10\%$ represents the lower

bound of the 90-110% range. Thus, the statistical analyses employed should not reject the null hypothesis if the difference is under 10%, but also maintaining an overall Type I error of $\alpha = .05$.

Benchmarks construction. Benchmarks have been created with as little as one RCT (Lee et al., 2013), two RCTs (Merrill, Tolbert, & Wade, 2003), three RCTs (Curtis et al., 2009; Reese et al., 2014), or several RCTs (Weersing & Weisz, 2002). Two sets of benchmarks were constructed for this study: (a) depression benchmarks and (b) client feedback benchmarks. Each set of benchmarks include an efficacy benchmark from the pre-post treatment outcomes of RCT treatment groups and a comparison benchmark from either the pre-post scores of the RCT waitlist/no-treatment control groups (for the depression benchmarks) or treatment-as-usual (TAU, for the client feedback benchmarks). Once clinical trials are selected, they are combined using standard meta-analytic procedures (e.g., Becker, 1988; Hedges & Olkin, 1985).

Depression benchmarks. After a systematic search of the extant clinical trial literature, I was unable to find RCTs meeting criteria—most importantly RCTs utilizing equivalent measures and intent-to-treat samples with youth diagnosed with depression. Given this paucity of suitable RCTs, I utilized Weiss et al.'s (2006) meta-analysis and identified RCTs to construct benchmarks for this study. Next, I will briefly describe the systematic literature search.

Clinical trial selection. Clinical trial studies were reviewed for eligibility in the construction of the benchmarks. I performed a systematic search of the literature borrowing criteria from Weisz et al.'s (2006) meta-analysis of psychotherapy with youth diagnosed with depression. The current literature search began where their search

stopped. Updating their meta-analysis required the following computer search strategy: (a) computer database searches (2005-present) on PsycINFO, Dissertation Abstracts International, and MEDLINE; and (b) reference list examination from outcome studies and relevant review articles. As part of the computer database searches, the keywords included *depression*, *dysthymia*, and *major depression*, and the search was limited to child and adolescent populations and studies that were treatment outcome, clinical trial, single-blind design, or double-blind design.

Continuing with Weisz et al.'s (2006) study selection criteria, the following criteria were applied: (a) a psychotherapy intervention was intended to target depressive symptoms or disorder; and (b) mean age of sample was younger than 18 years. In order to be most suitable for the current benchmarking study, two additional inclusion criteria were applied: (a) included an intent-to-treat sample since intent-to-treat samples are most comparable to effectiveness studies with comorbidity and premature drop-outs, and (b) included at least one psychotherapy outcome measure that has low reactivity (i.e., self-report) and low specificity (i.e., measures broad symptoms or global functioning) to have outcome measure equivalency with the ORS/CORS. Although some earlier benchmark studies (Weersing & Weisz, 2002) did not consider reactivity and specificity of outcome measures, more recent benchmarking studies (e.g., Minami, Wampold et al., 2008; Reese et al., 2014) have made a “best effort of equivalence” (Minami, Serlin et al., 2008, p. 517) by matching outcome measures by reactivity and specificity as much as possible.

The aforementioned systematic search in MEDLINE, PsychINFO, and Dissertation Abstracts International yielded no RCTs for benchmark construction. After excluding studies not meeting criteria—studies of adults (e.g., Cornelius et al., 2010,

Eskin et al., 2008), children too young (e.g., Cheng et al., 2007), medication-only (e.g., Cheung et al., 2008), non-randomization (e.g., Bahar et al., 2013; Melvin et al., 2013)—25 studies remained. I searched these 25 studies for inclusion of low reactivity-low sensitivity measures (like the ORS and CORS) used with youth in RCTs being treated for depression and intent-to-treat analyses. Only one study (Vitiello et al., 2006) met these criteria, albeit with a placebo control group.

Given this lack of suitable RCTs, I made a second attempt utilizing a recent article (Deighton et al., 2014) identifying 11 low-reactivity, low-specificity treatment outcome measures for youth in psychotherapy. I searched PsychINFO for RCTs using each of the 11 LR-LS measures listed by Deighton et al (2014) as a search term. Those 11 searches resulted in zero RCT studies meeting inclusion criteria.

Use of existing meta-analysis. As an alternative, I reviewed and selected all clinical trials from Weisz et al.'s (2006) meta-analysis that reported means and standard deviations for intent-to-treat and waitlist/no-treatment control group samples. Weisz et al's meta-analysis was chosen in part due to its rigorous methodology, including unpublished RCTs from dissertation research. First, thirteen RCTs (see Table 2; Clarke et al., 2001; Dana, 1998; De Cuyper, Timbremont, Braet, De Backer, & Wullaert, 2004; Diamond, Reis, Diamond, Siqueland, & Isaacs, 2002; Ettelson, 2003; Fischer, 1995;

Table 2

Treatment groups for Intent-To-Treat Depression Benchmark

Study	<i>N</i>	Treatment type	Age (<i>M</i> or range) in years	% Male	% White	<i>M</i> # of sessions	Measure(s)	Pretest		Posttest		<i>d</i>
								<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Clarke et al. (2001)	45	Cognitive restructuring	14.4	20	82	9.5	CES-D	25.2	8.70	17.8	8.70	0.83
Dana (1998)	10	CBT	8-13	NA	NA	8	CDI	23.0	11.80	14.3	8.41	0.67
De Cuyper et al. (2004)	9	CBT	10	NA	100	16	CDI	12.67	6.0	10.11	6.03	0.39
Diamond et al. (2002)	16	ABFT	13-17	NA	NA	8	BDI, YSR (INT)	23.8 65.4	7.4 9.3	10.4 56.4	9.8 10.8	1.32 0.92
Ettelson (2003)	13	CBT	9-12 grade	46	100	16	CDI ^a	69.31	15.47	62.08	12.16	0.44
Fischer (1995)	8	CBT	12-17	NA	100	5	BDI	24.25	10.34	17.75	9.15	0.56
Kahn et al. (1990)	17	CBT	10-14	NA	NA	12	RADS, CDI	87.00 31.11	8.96 9.58	53.44 7.29	14.72 66.03	2.97 2.37
Kahn et al. (1990)	17	Relaxation	10-14	NA	NA	12	RADS, CDI	83.44 26.94	8.03 10.83	61.76 12.88	14.86 10.71	1.90 1.24
Kahn et al. (1990)	17	Self- modeling	10- 14	NA	NA	NA	RADS, CDI	81.65 27.18	10.66 7.84	62.12 13.58	12.00 7.38	1.70 1.65
Mufson et al. (1999)	24	IPT	15.7	25	NA	12	BDI	18.8	8.50	5.9	8.10	1.47

Table 2 continued

Mufson et al. (2004)	34	IPT	15.1	16	NA	12	BDI	20.8	8.70	8.4	11.0	1.39
Rohde et al. (2004)	44	CBT	15.1	40	80	8.4	BDI-II	16.6	12.80	9.6	10.70	0.54
TADS (2004)	111	CBT	12-17	NA	NA	11	RADS	78.69	10.59	67.96	14.18	1.01
Vostanis et al. (1996)	29	CBT	8-17	NA	NA	9	Mood & Feelings Questionnaire	33.4	12.20	17.6	15.2	1.26
Weisz et al. (1997)	16	CBT	9.4	56	75	8	CDI	18.63	5.32	7.06	6.12	2.06

Notes. Studies selected from Weisz et al. (2006); N = sample size; $d = [1 - (3/4n - 5)] [M_{\text{post}} - M_{\text{pre}}/SD_{\text{pre}}]$; M = Mean; SD = Standard Deviation; NA = not available; CBT = Cognitive Behavioral Therapy; IPT = Interpersonal Process Therapy; ABFT = Attachment-based family treatment; CES-D = Center for Epidemiologic Studies Depression; BDI = Beck Depression Inventory; BDI-II = Beck Depression Inventory-II; RADS = Reynolds Adolescent Depression Scale; CDI = Children's Depression Inventory; Treatment for Adolescents with Depression Study Team; YSR (INT) = Youth Self-Report (internalizing subscale)

^a = Total T-Score reported

Table 3

Control Groups for Waitlist Depression Benchmark

Study	<i>N</i>	Condition	Age (<i>M</i>	%	%	<i>Time</i>	Measure(s)	Pretest		Posttest		<i>d</i>
			or range) in years	Mal e	Whit e			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Ackerson et al. (1998)	10	Waitlist	15.9	30	60	4 weeks	CDI	16.8	4.5	15.8	5.2	0.20
Clarke et al. (1999)	27	Waitlist	16.2	NA	NA	8 weeks	BDI	24.2	10.8	16.0	11.2	0.74
Clarke et al. (1995)	68	NT	NA	NA	NA	5 weeks	CES-D	21.88	9.2	21.67	12.3	0.02
Curtis (1992)	9	Waitlist	16.1	0	0.89	8 weeks	BDI, RADS	24.6 61.0	6.4 3.8	22.5 56.7	10.5 14.2	0.66 1.02
Dana (1998)	9	NT	8-13	NA	NA	8 weeks	CDI	21.0	14.11	13.33	9.45	0.49
De Cuyper et al. (2004)	10	Waitlist	9-11	NA	100	16 weeks	CDI	15.27	4.54	11.73	5.66	0.71
Diamond et al. (2002)	16	Waitlist	13-17	NA	NA	6 weeks	BDI, YSR (INT)	28.0 66.6	7.1 6.6	18.5 61.9	11.1 8.4	0.97 0.68
Ettelson (2003)	12	Waitlist	9-12 grade	42	92	8 weeks	CDI ^a	63.25	11.25	71.33	17.73	-0.67
Kahn (1989)	9	Waitlist	16.45	80	100	6 weeks	BDI	10.5	5.505	5.77	4.63	0.78

Table 3 continued

Kahn (1989)	6	Waitlist	15.33	50	83	6 weeks	BDI	9.08	4.49	6.33	3.01	0.52
Kahn et al. (1990)	1	Waitlist	10-14	NA	NA	8 weeks	RADS,	85.41	10.98	80.12	13.44	0.28
	7						CDI	28.06	9.75	26.94	15.41	0.11
Lewinsohn et al. (1990)	1	Waitlist	16.28	32	NA	7-8	CES,	14.89	4.3	12.89	4.74	0.36
	9					weeks	BDI	23.84	11.43	20.47	10.28	0.28
Liddle & Spence (1990)	1	NT	7-12	NA	NA	8 weeks	CDI	20.7	3.34	16.9	6.79	1.04
	0											
Marcotte & Baron (1993)	1	Waitlist	14-17	23	NA	6 weeks	BDI	21.39	6.33	14.85	7.13	0.97
	3											
Reynolds & Coats (1986)	1	Waitlist	9-12	NA	100	10 weeks	BDI,	16.9	5.48	18.31	9.82	-0.17
	0		Grade				RADS	80.7	3.58	81.12	13.46	0.11
Rosello & Bernal (1999)	1	Waitlist	13-17	NA	NA	12 weeks	CDI	20.13	5.99	15.83	6.83	0.69
	8											
Stark et al. (1987)	9	Waitlist	11.3	56	NA	5 weeks	CDI,	20.11	9.88	18.6	9.91	0.23
							CDS	67.56	17.8	61.09	16.73	0.33
Weisz et al. (1997)	3	NT	9.7	53	56	8 weeks	CDI	17.81	10.05	11.81	10.0	0.58
	2											

Notes: Studies selected from Weisz et al. (2006); *N* = sample size; NT = No Treatment; NA = not available; $d = [1 - (3/4n - 5)] [M_{\text{post}} - M_{\text{pre}}/SD_{\text{pre}}]$; *M* = Mean; *SD* = Standard Deviation; CBT = Cognitive Behavioral Therapy; IPT = Interpersonal Process Therapy; CES-D = Center for Epidemiologic Studies Depression; BDI = Beck Depression Inventory; RADS = Reynolds Adolescent Depression Scale; CDI = Children's Depression Inventory; YSR (INT) = Youth Self-Report (Internalizing scale).

^a = Total T-Score reported.

Kahn, Kehle, Jenson, & Clark, 1990; Mufson et al., 2004; Mufson, Weissman, Moreau, & Garfinkel, 1999; Rohde, Clarke, Mace, Jorgensen, & Seeley, 2004; Treatment for Adolescents With Depression Study (TADS) Team, 2004; Vostanis, Feehan, Grattan, & Bickerton, 1996; Weisz, Thurber, Sweeney, Proffitt, & LeGagnoux, 1997) reported information for intent-to-treat treatment groups and were included in the intent-to-treat depression treatment efficacy benchmark. Second, seventeen RCTs (see Table 3; Ackerson, Scogin, McKendree-Smith, & Lyman, 1998; Clarke et al., 1995; Clarke, Rohde, Lewinsohn, Hops, & Seeley, 1999; Curtis, 1992; Dana, 1998; De Cuyper et al., 2004; Diamond et al., 2002; Ettelson, 2003; Kahn, 1989; Kahn et al., 1990; Lewinsohn, Clarke, Hops, & Andrews, 1990; Liddle & Spence, 1990; Marcotte & Baron, 1993; Reynolds & Coats, 1986; Rosello & Bernal, 1999; Stark, Reynolds, & Kaslow, 1987; Weisz et al., 1997) reported means and standard deviations of waitlist/no treatment groups and these samples were utilized in calculating a waitlist control efficacy benchmark ES.

Client feedback (complete sample) benchmarks. The second set of benchmarks was also derived from the best effort of equivalence. No extant literature permitted a benchmark with client feedback with youth. Only one RCT has been conducted for routine client feedback with youth (Bickman et al., 2011), but several considerations do not allow for comparison: (a) they did not report means and standard deviations for their samples, (b) one third of the clinicians in the treatment condition did not utilize client feedback, and (c) feedback was not available to clinicians until nine days after reported by clients. Given this paucity of client feedback RCTs with youth, we utilized recent

client feedback benchmarks constructed for a benchmarking study for adults in a public behavioral health setting (Reese et al., 2014).

Reese and colleagues (2014) performed a systematic review of the literature to find RCTs for benchmark construction. They also utilized previous client feedback meta-analyses (Lambert & Shimokawa, 2011; Shimokawa et al., 2010). Their search resulted in nine studies: six studies using the Outcome Questionnaire System (Lambert, 2010), and three studies using PCOMS (Anker et al., 2009; Reese et al., 2009). See Lambert & Shimokawa (2011) for a thorough review of the studies. Then, they used benchmarking formulas outlines by Minami, Serlin, et al. (2008) to compute unbiased ESs (same formula as used in this study) and to aggregate these ESs across the feedback studies. They reported four benchmark ESs, which I will use for benchmarking purposes: (a) feedback treatment condition ES from all nine studies ($d_{FTall} = 0.60$), (b) TAU condition ES from all nine studies ($d_{TAUall} = 0.41$), (c) feedback treatment condition ES from three PCOMS studies ($d_{FTors} = 1.13$), and (d) TAU condition ES from three PCOMS studies ($d_{TAUors} = 0.47$). I will utilize these ESs for benchmarking purposes with the current full PBH sample ($N = 4,389$) regardless of pretreatment scores or diagnoses.

Depression efficacy trial benchmark effect size calculations. Next, the efficacy trial depression benchmarks were calculated for studies from Weisz et al. (2006). In keeping with benchmarking methodology, I only included pre- and posttest results from self-report outcome measures related to the primary diagnosis or dependent variable in the study. When means and standard deviations were available for two self-report measures within a study, effect sizes were calculated separately for each measure and then aggregated using the mean of the ESs to obtain a single pre-post ES for the waitlist

control group and intent-to-treat group (Weisz et al., 2006). The formula ($d_i = [1 - (3/(4n - 5))] [(M_{\text{post}} - M_{\text{pre}})/SD_{\text{pre}}]$) for calculating an unbiased Cohen's d effect size was employed consistent with recent benchmarking studies (Minami, Serlin, et al., 2008; Reese et al., 2014) where n is the sample size, SD_{pre} is the pretreatment standard deviation, and M_{pre} and M_{post} are the pre- and post-treatment means. Effect size estimates are particularly important to demonstrate practical importance and guard against Type I errors (e.g., clinically trivial differences may be due to statistical power with a large sample size).

After effect sizes for each study were calculated, they were aggregated across clinical trials to yield single ESs, serving as comparison benchmarks. I combined the ESs into an aggregated benchmark ES following meta-analytic procedures outlined by Minami, Serlin, et al. (2008). Specifically, the variance of each RCT d_i is estimated by

$$\hat{\sigma}_{d(i)}^2 = \frac{2(1-r_i)}{n_i} + \frac{d_i^2}{2n_i}$$

with r_i being the estimated correlation between the pre- and post-treatment scores of the outcome measure (Becker, 1988). Consistent with previous benchmarking studies of treatment outcome for depression (e.g., Minami et al., 2007, Reese et al., 2014), a reasonable estimate for outcome measures of depression treatment is $r = 0.5$.

All ESs were aggregated into a benchmark ES using

$$d_{\text{WL}} = \sum_i \frac{d_i}{\hat{\sigma}_{d(i)}^2} / \sum_i \frac{1}{\hat{\sigma}_{d(i)}^2}$$

for the efficacy benchmark waitlist condition ES, and

$$d_{\text{ITT}} = \sum_i \frac{d_i}{\hat{\sigma}_{d(i)}^2} / \sum_i \frac{1}{\hat{\sigma}_{d(i)}^2}$$

for the benchmark treatment condition ES.

The aggregation resulted in a waitlist control benchmark ES of $d_{WL} = 0.37$ and an intent-to-treat treatment group benchmark ES of $d_{ITT} = 1.01$.

Critical value calculation. Calculating critical values allows for statistical testing with range-null hypotheses while maintaining an overall Type I error rate of $\alpha = .05$, thereby permitting reasonable conclusions about comparability (Serlin & Lapsley, 1985, 1993). Following Minami, Serlin, et al. (2008), the benchmarking hypotheses rely on a 95th percentile test statistic, (e.g., $t_{(ITT)\nu, \lambda: .95}$ and $t_{(WL)\nu, \lambda: .95}$), which follows a noncentral t distribution ($\nu = N - 1$ degrees of freedom) and has a noncentrality parameter $\lambda = \sqrt{N}(d_{ITT} - 10\%)$ or $\lambda = \sqrt{N}(d_{WL} + 10\%)$. The critical values for the depression-related benchmarks were determined by a normal approximation of the distribution and resulted in ($d_{CV(ITT)} = 1.00$) and ($d_{CV(WL)} = 0.49$).

These critical values allow for range-null hypotheses, maintaining an overall Type I error rate of $\alpha = .05$. Serlin and Lapsley (1985, 1993) succinctly state: "...if the critical value is chosen so that the Type I error rate equals α when λ is at the limit allowed under H_0 , then because all other values of λ under the null hypothesis are smaller than this upper [or lower] limit, the Type I error rate under H_0 is guaranteed to be at most α ."

Similarly, I calculated critical values for the client feedback benchmarks based on the ESs reported by Reese et al. (2014) but in relation to the current PBH full sample size ($N = 4,389$). Specifically, the critical values from all nine client feedback studies were $d_{cv} = 0.57$ for the feedback treatment condition and $d_{cv} = 0.48$ for the TAU condition. Similarly, the critical values based on the three ORS client feedback studies were $d_{cv} = 1.05$ for the feedback treatment condition and $d_{cv} = 0.54$ for the TAU condition. In other words, the two feedback treatment condition ESs have their corresponding critical values

associated with the lower bound of the 10 percent range of clinical equivalence, and the two TAU condition ESs have their corresponding critical values associated with the upper bound of the 10 percent range of clinical equivalence.

Data Analysis

Means, standard deviations, and ESs were calculated for the full PBH sample by study sample characteristics, including age, sex, ethnicity (see Table 4), and primary diagnoses (see Table 5).

Effect size calculations. Next, pre-post ESs (Cohen's d) were calculated for the full PBH sample and depression sample using baseline (pre-counseling), endpoint (post-counseling), and standard deviations from client ORS/CORS scores. Consistent with ES calculation for clinical trials noted earlier, I used the formula ($d_i = [1 - (3/(4n - 5))] [(M_{\text{post}} - M_{\text{pre}})/SD_{\text{pre}}]$) to calculate unbiased effect size estimates for the full PBH sample and PBH depression sample, where n is the sample size, M is the mean of the measure, and SD is the standard deviation. These ESs allowed comparison with the previously published efficacy studies contained in the benchmark ESs. Variances of the current PBH samples effect sizes d_{PBHcf} and d_{PBHdep} were also estimated and reported using

$$\hat{\sigma}_{d(i)}^2 = \frac{2(1-r_i)}{n_i} + \frac{d_i^2}{2n_i}.$$

Here again, r_i is the estimated correlation between the pre- and post-treatment scores of the outcome measure (Becker, 1988). The Pearson product-moment correlation coefficients were calculated for the PBH's full sample ($r = 0.326$) and depression sample ($r = 0.305$).

Benchmarking analyses. Finally, benchmarking analyses were conducted. In order to adequately compare the effect size of the PBH samples to the selected clinical trial benchmarks, range null hypotheses were employed since the large sample sizes would likely lead to a false rejection of a conventional point null hypothesis. Therefore, a range null hypothesis with an *a priori* 10% margin of difference between the benchmark and PBH ESs was employed. As mentioned previously, a range-null hypothesis follows a noncentral t statistic (Serlin & Lapsley, 1985, 1993) and a normal distribution is approximated. In order to perform statistical testing, critical values (previously described in detail) associated with the different efficacy trial benchmarks were employed representing the 95th percentile value of the noncentral t distribution. Hence, the PBH effect size estimate was evaluated through (a) clinical significance testing using the 10% margin of difference surrounding the efficacy trial benchmark and (b) statistical testing maintaining an overall Type I error rate of $\alpha = .05$, permitting reasonable conclusions about comparability.

Benchmarking against treatment groups. For the PBH depression sample effect size d_{PBHdep} to be considered clinically equivalent to the intent-to-treat treatment efficacy trial benchmark d_{itt} , the PBH effect size needs to exceed the critical value

$$d_{CVitt} = \frac{t_{v,\lambda:95}}{\sqrt{N}},$$

where $t_{v,\lambda:95}$ is the 95th percentile value of the noncentral t distribution and $\lambda = \sqrt{N}(d_{itt} - 10\%)$ is the noncentrality parameter.

Similarly, for the PBH full sample effect size d_{PBHcf} to be considered clinically equivalent to the client feedback efficacy trial benchmarks d_{FTall} or d_{FTors} , the PBH effect size needs to exceed the critical value

$$d_{CVall} = \frac{t_{v,\lambda:95}}{\sqrt{N}}$$

or

$$d_{CVors} = \frac{t_{v,\lambda:95}}{\sqrt{N}}$$

respectively. For each, $t_{v,\lambda:95}$ is the 95th percentile value of the noncentral t distribution and $\lambda = \sqrt{N}(d_i - 10\%)$ is the noncentrality parameter.

Benchmarking against waitlist control and treatment-as-usual conditions.

Significance testing for comparison to the waitlist control depression benchmark and TAU full sample client feedback benchmarks is similar to the above treatment efficacy benchmarks but with +10% in the formula replacing -10% for the noncentrality parameter. In other words, if the PBH effect size estimate for depression-related treatment does not statistically significantly exceed the waitlist benchmark at the 10 percent critical value, treatment is considered practically and statistically equivalent to a waitlist or no treatment control condition (change rates observed in natural remission of psychological distress). Similarly, if the PBH effect size estimate for client feedback treatment for the full sample does not statistically significantly exceed the TAU benchmark at the 10 percent critical value, treatment is considered practically and statistically equivalent to a TAU condition in client feedback clinical trials.

Research Hypotheses

I have four hypotheses in this study. For the first two, I hypothesize that the treatment outcomes for the current PBH full sample will be (a) statistically and clinically equivalent to efficacy outcomes of the feedback treatment condition observed in the client feedback RCTs and (b) statistically and clinically superior to TAU conditions from the same client feedback RCTs.

For the second two, I hypothesize that the treatment outcomes for the PBH depression sample will be (a) statistically and clinically equivalent to efficacy outcomes of the treatment efficacy condition observed in the intent-to-treat (ITT) analysis of RCTs and (b) statistically and clinically superior to wait list (WL) control conditions from RCTs.

Hypothesis one. Following range-null hypothesis testing guidelines from Serlin and Lapsley (1985, 1993) and exemplified by Minami, Serlin, et al. (2008), when δ_{PBHcf} is the true population effect size estimate of the PBH treatment (in Cohen's d), δ_{FTall} and δ_{FTors} are the true population treatment client feedback efficacy benchmarks (in Cohen's d), and 10% is the maximum difference allowed to claim clinical equivalence, the range-null and alternative hypotheses are:

$$H_0 : \delta_{PBHcf} \leq \delta_{FTall} - 10\%$$

$$H_1 : \delta_{PBHcf} > \delta_{FTall} - 10\%$$

and

$$H_0 : \delta_{PBHcf} \leq \delta_{FTors} - 10\%$$

$$H_1 : \delta_{PBHcf} > \delta_{FTors} - 10\%.$$

Hypothesis two. If the PBH full sample effect size estimate does not exceed the TAU comparison benchmarks by 110% it will be deemed clinically comparable to the TAU benchmarks. So, when δ_{PBHcf} is the true population effect size estimate of the PBH treatment (in Cohen's d), δ_{TAUall} and δ_{TAUors} are the true population treatment TAU efficacy benchmarks (in Cohen's d), and 10% is the maximum difference allowed to claim clinical equivalence, the range-null and alternative hypotheses are:

$$H_0 : \delta_{PBHcf} \leq \delta_{TAUall} + 10\%$$

$$H_1 : \delta_{PBHcf} > \delta_{TAUall} + 10\%$$

and

$$H_0 : \delta_{PBHcf} \leq \delta_{TAUors} + 10\%$$

$$H_1 : \delta_{PBHcf} > \delta_{TAUors} + 10\%.$$

Hypothesis three. Again, following guidelines from Serlin and Lapsley (1985, 1993) and exemplified by Minami, Serlin, et al. (2008), when δ_{PBHdep} is the true population effect size estimate of the PBH treatment (in Cohen's d), for the depression sample, δ_{ITT} is the true population intent-to-treat efficacy benchmark (in Cohen's d), and 10% is the maximum difference allowed to claim clinical equivalence, the range-null and alternative hypotheses are:

$$H_0 : \delta_{PBHdep} \leq \delta_{ITT} - 10\%$$

$$H_1 : \delta_{PBHdep} > \delta_{ITT} - 10\%$$

Hypothesis four. If the PBH depression sample effect size estimate does not exceed the waitlist control benchmark by 110% it will be deemed clinically comparable to the waitlist control benchmark. So, when δ_{PBHdep} is the true population effect size estimate of the PBH treatment (in Cohen's d), δ_{WL} is the true population waitlist efficacy benchmark (in Cohen's d), and 10% is the maximum difference allowed to claim clinical equivalence, the range-null and alternative hypotheses are:

$$H_0 : \delta_{PBHdep} \leq \delta_{WL} + 10\%$$

$$H_1 : \delta_{PBHdep} > \delta_{WL} + 10\%.$$

Chapter Three: Results

Preliminary Analysis

First, I calculated average session numbers. The average number of sessions for the PBH full sample ($N = 4,389$) was 12.34, $SD = 14.366$. The average number of sessions for the PBH depression sample ($N = 469$) was 11.67, $SD = 11.352$.

Table 4

Full Sample Therapy Outcomes by Client Demographics

	Sample Size	Pre ORS $M (SD)$	Post ORS $M (SD)$	Within Group $d (95\% CI)$
6- to 12- year-olds	2,244	27.09 (7.56)	33.04 (7.09)	0.79 [0.73, 0.85]
13- to -17-year-olds	2,124	25.76 (7.38)	30.97 (7.25)	0.71 [0.66, 0.76]
Female n (%)	2,165	25.73 (7.51)	31.47 (7.41)	0.76 [0.71, 0.81]
Male n (%)	2,207	27.13 (7.43)	32.57 (7.04)	0.73 [0.68, 0.78]
Latino(a)/Hispanic n (%)	1,002	26.48 (7.71)	32.66 (7.06)	0.80 [0.72, 0.88]
African American n (%)	309	27.14 (7.62)	30.90 (8.00)	0.49 [0.36, 0.62]
Native American n (%)	55	26.79 (6.90)	32.46 (6.93)	0.81 [0.46, 1.16]
Euro-American n (%)	1,416	26.60 (7.49)	31.43 (7.08)	0.65 [0.59, 0.71]
Other Ethnicity n (%)	102	27.23 (8.22)	32.74 (6.14)	0.67 [0.42, 0.92]
Unknown Ethnic n (%)	1,505	26.05 (7.30)	32.31 (7.39)	0.86 [0.79, 0.93]

Notes. $d = [1 - (3/4n - 5)] [M_{\text{post}} - M_{\text{pre}}/SD_{\text{pre}}]$. C/ORS = Outcome Rating Scale or Child Outcome Rating Scale; CI = confidence interval

I also tested the current PBH full sample for disparities in treatment outcomes based on client age group, sex, ethnicity, and diagnosis. First, I tested whether children (i.e., 6- to 12-year-olds) utilizing the CORS and adolescents (i.e., 13- to 17-year-olds)

utilizing the ORS had similar outcomes via an ANOVA. Age category was the independent variable (IV) and C/ORS pre-post change score was the dependent variable (DV). The result for age category was statistically significant, $F(1, 4,388) = 8.16, p = .004$, partial $\eta^2 = .002$, with children ($n = 2,244$) having larger pre-post change scores than adolescents (see Table 4 for effect size estimates). Despite the statistical significance found, a partial η^2 of .002 means that only 0.2% of the variance was explained by whether the client was a child or adolescent. This indicates that statistical significance does not have any practical application and was likely due to a large sample size (i.e., Type I error). Since age category is synonymous with whether youth completed the ORS or CORS, the finding that only 0.2% of the variance is explained by this difference provided reasonable justification for analyzing the ORS and CORS scores together rather than separately. Additionally, mean pre-post change scores for ORS ($M = 5.21$) and CORS ($M = 5.95$) groups were minimally different (0.74).

Second, I tested differences in treatment outcome by sex of the client. Sex category (i.e., male or female), was the IV and C/ORS pre-post change score was the DV. Treatment outcomes differences by sex category were not statistically significant, $F(1, 4,370) = 1.32, p = .25$, partial $\eta^2 < .001$.

Third, I tested whether clients of different racial/ethnic categories had similar treatment outcomes. Race/ethnicity category (for all known race/ethnicities) was the IV and C/ORS pre-post change score was the DV. Treatment outcome differences by race/ethnicity were statistically significant, $F(4, 2,879) = 6.39, p < .001$, partial $\eta^2 = .009$. For each race/ethnicity, sex and age group, effect sizes were estimated separately (see Table 4). A partial η^2 of .009 means that only 0.9% of the variance was explained by

race/ethnic category of the client, and a finding of statistical significance was likely due to a large sample size (i.e., Type I error).

I conducted post-hoc analyses between all racial/ethnic groups to determine the exact nature of the differences. Significance-testing was based on Bonferroni corrections for multiple analyses ($\alpha = .0083$) for six comparisons between the four known racial/ethnic categories. Two statistically significant differences were found: Youth identifying as Latino(a)/Hispanic had significantly larger mean pre-post change scores than African American youth ($M = 2.4293$, $SE = 0.5461$, $p < .001$) and Euro-American youth ($M = 1.3303$, $SE = 0.3464$, $p = .001$). Interactions between age (adolescent or child) and ethnic group were not significant for mean pre-post change scores $F(3, 2774) = 1.84$, $p = .138$, partial $\eta^2 = .002$.

Finally, I tested via ANOVA whether treatment outcomes differed by primary diagnosis, with primary diagnosis as the IV and ORS pre-post change score as the DV. Treatment outcome differences by diagnosis were statistically significant, $F(13, 4,119) = 2.74$, $p = .001$, partial $\eta^2 = .009$. A partial η^2 of .009 means that only 0.9% of the variance was explained by the primary diagnosis, and a finding of statistical significance was likely a false positive (i.e., Type I error) due to a large sample size. Given 14 diagnostic categories, post-hoc analysis between categories was untenable with Bonferroni correction. For the most frequently reported diagnostic categories (i.e., $n < 25$), effect sizes were estimated separately and reported in Table 5 for the full sample and Table 6 for the depression-related clinical sample.

Table 5

Full Sample Therapy Outcomes by Diagnosis

	Sample Size	Pre C/ORS <i>M</i> (<i>SD</i>)	Post C/ORS <i>M</i> (<i>SD</i>)	Within Group <i>d</i> (95% <i>CI</i>)
Substance Dx	91	26.68 (6.92)	30.70 (8.29)	0.58 [0.34, 0.82]
Major Depression	88	24.19 (7.15)	29.43 (8.59)	0.73 [0.46, 1.00]
Bipolar Dx	145	24.45 (7.39)	29.58 (8.03)	0.69 [0.48, 0.90]
Mood Disorder NOS	429	25.13 (8.07)	30.37 (8.15)	0.65 [0.53, 0.77]
Anxiety Dx	209	26.42 (7.16)	31.12 (7.30)	0.65 [0.48, 0.82]
Adjustment Dx	1,189	26.81 (7.29)	33.41 (6.56)	0.90 [0.82, 0.98]
PTSD	82	24.32 (8.70)	29.12 (8.48)	0.55 [0.29, 0.81]
Physical Abuse of Child	72	27.10 (7.47)	32.01 (7.56)	0.65 [0.37, 0.93]
Neglect of Child	29	28.43 (6.02)	33.19 (6.65)	0.77 [0.29, 1.25]
Disruptive Behavior Dx	213	25.90 (7.64)	31.90 (7.52)	0.78 [0.60, 0.96]
Depression Disorder NOS	197	23.37 (7.85)	30.10 (7.64)	0.85 [0.69, 1.01]
ADHD	575	27.33 (7.61)	32.34 (6.82)	0.66 [0.55, 0.77]
Pervasive Dev Disorders	43	27.58 (5.84)	33.56 (7.20)	1.00 [0.57, 1.43]
V-codes	771	27.28 (7.10)	32.21 (6.79)	0.69 [0.60, 0.78]

Notes. $N = 4,133$; $d = [1 - (3/4n - 5)] [M_{\text{post}} - M_{\text{pre}}/SD_{\text{pre}}]$; M = Mean; SD = Standard Deviation; Dx = Diagnosis; C/ORS = Outcome Rating Scale or Child Outcome Rating Scale; CI = confidence interval; NOS = not otherwise specified; Diagnoses reflect the primary diagnosis. Anxiety Dx = diagnosis of panic, panic w/ and w/o agoraphobia, anxiety NOS, phobias, obsessive-compulsive disorder, or generalized anxiety disorder; Adjustment Dx = any adjustment diagnosis; PTSD = posttraumatic stress disorder; ADHD = attention-deficit/hyperactivity disorder; Disruptive Behavior Disorders = Conduct Disorder, Oppositional Defiant Disorder, Oppositional Defiant Disorder, Disruptive Behavior Disorder NOS; Pervasive Dev Disorders = Autistic Disorder, Asperger's Disorder, and Pervasive Developmental Disorder NOS; Missing data due to diagnosis not reported or infrequent diagnoses ($n < 25$; e.g., schizophrenia, eating disorders, psychotic disorder NOS, dysthymia, and deferred diagnoses).

Table 6

Depression-related Clinical Sample Therapy Outcomes by Diagnosis

	Sample Size	Pre C/ORS <i>M</i> (<i>SD</i>)	Post C/ORS <i>M</i> (<i>SD</i>)	Within Group Effect Size <i>d</i> (95% CI)
Major Depression	60	20.83 (5.49)	28.51 (8.96)	1.38 [1.37, 1.65]
Depression NOS	134	19.46 (6.20)	28.13 (7.81)	1.39 [1.14, 1.64]
Dysthymic Dx	17	21.77 (5.51)	29.65 (4.80)	1.36 [0.58, 2.14]
Adj Dx w/ Dep	163	20.80 (6.60)	31.15 (7.86)	1.56 [1.31, 1.81]
Adj Dx w/ Mixed	95	22.25 (5.31)	31.64 (6.02)	1.75 [1.39, 2.11]

Notes. $N = 469$; $d = [1 - (3/4n - 5)] [M_{\text{post}} - M_{\text{pre}}/SD_{\text{pre}}]$; M = Mean; SD = Standard Deviation; Dx = Diagnosis; C/ORS = Outcome Rating Scale or Child Outcome Rating Scale; CI = confidence interval; NOS = not otherwise specified; Adj = Adjustment; w/ Mixed = with mixed anxiety and depression; w/ Dep = with Depressed Mood; Diagnoses reflect the primary diagnosis.

Results of Client Feedback Benchmark Hypotheses

The following results are from research hypotheses related to the PBH full sample. The mean pre-post treatment ORS/CORS scores for the PBH full sample ($N = 4,389$) were $M_{\text{pre}} = 26.44$ ($SD = 7.50$) and $M_{\text{post}} = 32.02$ ($SD = 7.25$), respectively, resulting in an observed standardized pre-post mean change of $d_{\text{PBHcf}} = 0.74$ (see Table 7) with a variance of 0.0004. All analyses utilized critical values with a Type I error rate of $\alpha = 0.05$.

Table 7

Effect Size Comparisons to Client Feedback Benchmark RCT Studies

PBH d	Feedback benchmark (all)		Feedback benchmark (ORS)		TAU benchmark (all)		TAU benchmark (ORS)	
	d_{cv}	p	d_{cv}	p	d_{cv}	p	d_{cv}	p
0.74	0.57	<.001	1.05	>.999	0.48	<.001	0.54	<.001

Notes. PBH = Public Behavioral Health; TAU = treatment as usual; ORS = Outcome Rating Scale; RCT = randomized clinical trial; $d = [1 - (3/4n - 5)] [M_{post} - M_{pre}/SD_{pre}]$; d_{cv} = critical effect size value required to attain statistical significance.

Hypothesis one. My first hypothesis was that the effect size estimate for the full PBH sample ($d_{PBHcf} = 0.74$; $N = 4,389$) would be clinically equivalent to the treatment benchmark made up of feedback treatment outcomes, as reported in Reese et al. (2014) for all nine client feedback clinical trials and separately for the three ORS clinical trials. First, compared against the full feedback treatment benchmark ($d_{FTall} = 0.60$) with a 10% a priori margin ($d_{FTall} [90\%] = 0.54$) and critical value $d_{cv(FTall)} = 0.57$, the observed PBH effect size estimate was considered clinically equivalent ($t = 49.02$, $df = 4,388$, $\lambda = 35.77$, $p < .001$) to the feedback treatment group outcomes in the nine feedback studies. In other words, the estimated effect of the full PBH sample exceeded the critical value $d_{cv(FTall)} = 0.57$ —the magnitude of effect necessary to claim at least equivalence with feedback treatment outcomes from all nine clinical trials. Second, compared against the feedback treatment benchmark for the three ORS clinical trials ($d_{FTors} = 1.13$), given a 10% a priori margin ($d_{FTors} [90\%] = 1.02$) and critical value $d_{cv(FTors)} = 1.05$, the observed PBH effect size estimate did not achieve clinical equivalence ($t = 49.02$, $df = 4,388$, $\lambda = 67.38$, $p > .999$). In other words, the treatment outcomes of the full PBH sample cannot be

considered clinically equivalent to the feedback treatment group outcomes observed in the three ORS efficacy trials.

Hypothesis two. My second hypothesis was that the effect size estimate for the full PBH sample ($d_{PBHcf} = 0.74$; $N = 4,389$) would be clinically superior to TAU benchmarks from client feedback clinical trials: (a) $d_{TAUall} = 0.41$ for all nine client feedback studies, and (b) $d_{TAUors} = 0.47$ for the three ORS clinical trials. First, compared against the TAU benchmark ($d_{TAUall} = 0.41$), given a 10% a priori margin ($d_{TAUall} [110\%] = 0.45$) and critical value $d_{cv(TAUall)} = 0.48$ for all nine feedback treatment clinical trials, the observed PBH effect size estimate was statistically significant ($t = 49.02$, $df = 4,388$, $\lambda = 29.88$, $p < .001$), suggesting that treatment at the PBH was clinically superior to the TAU group outcomes in the nine feedback studies. Second, compared against the TAU benchmark ($d_{TAUors} = 0.47$), given a 10% a priori margin ($d_{TAUors} [110\%] = 0.52$) and critical value $d_{cv(TAUors)} = 0.54$ for the three ORS clinical trials, the observed PBH effect size estimate was also statistically significant ($t = 49.02$, $df = 4,388$, $\lambda = 34.25$, $p < .001$), suggesting that treatment at PBH also achieved clinical superiority to the TAU group outcomes found in the three ORS feedback studies.

Results of Depression Benchmark Hypotheses

The following results are from research hypotheses related to the PBH depression sample. The mean pre-post treatment ORS/CORS scores for the PBH depression sample ($N = 469$) were $M_{pre} = 20.75$ ($SD = 6.12$) and $M_{post} = 30.00$ ($SD = 7.69$), respectively, resulting in an effect size estimate of $d_{PBHdep} = 1.51$ (see Table 8) with a variance of 0.005. All analyses utilized critical values with a Type I error rate of $\alpha = 0.05$.

Table 8

Effect Size Comparisons to Depression Benchmark RCT Studies

PBH d	ITT benchmark		Waitlist control benchmark	
	d_{cv}	p	d_{cv}	p
1.51	1.00	<.001	0.49	.001

Notes. PBH = Public Behavioral Health; ITT = intent-to-treat; RCT = randomized clinical trial; $d = [1 - (3/4n - 5)] [M_{post} - M_{pre}/SD_{pre}]$; d_{cv} = critical effect size value required to attain statistical significance.

Hypothesis three. My third hypothesis was that the effect size estimate of the depression PBH sample ($N = 469$) would be clinically equivalent to the intent-to-treat efficacy benchmark from RCTs of youth in treatment for depression. Given the ITT benchmark ($d_{ITT} = 1.01$) with a 10% a priori margin ($d_{ITT} [90\%] = 0.90$) had an associated critical value of $d_{cv(ITT)} = 1.00$. The observed PBH effect size estimate ($d_{PBHdep} = 1.51$) was statistically significant ($t = 32.70$, $df = 469$, $\lambda = 19.68$, $p < .001$), and easily surpassed the critical value, suggesting at least clinical equivalence with ITT treatment efficacy from RCTs of youth in treatment for depression. Interestingly, seven of the 13 samples included in the ITT benchmark had 100% completion rates in their treatment samples, and others had low attrition rates.

Hypothesis four. My fourth hypothesis was that the effect size estimate of the depression PBH sample ($N = 469$) would be clinically superior to the waitlist control benchmark from RCTs of youth in treatment for depression. Compared against the waitlist control benchmark effect size estimate ($d_{WL} = 0.37$) given a 10% a priori margin ($d_{WL} [110\%] = 0.41$) and the associated critical value ($d_{cv(WL)} = 0.49$), the observed PBH

effect size estimate ($d_{PBHdep} = 1.51$) was also statistically significant ($t = 32.70$, $df = 469$, $\lambda = 8.81$, $p < .001$), and exceeded the associated critical value, suggesting that treatment at PBH was clinically superior to waitlist control conditions from RCTs of youth in treatment for depression.

Clinical Significance

Another way to understand clinical effectiveness is examining the rates of reliable change and clinically significant change. As previously mentioned, the reliable change index (RCI) on the ORS is any change of 6 points or more (Duncan, 2014). Also, the ORS has a clinical cut-off score of 28 for youth aged 13 to 17. Combining the RCI and clinical cutoff scores, the ORS can say that an adolescent initially scoring below 28 will experience reliable and clinically significant change by scoring 6 points higher and scoring at or above 28. Adolescents with pre-post changes of 6 points or greater but did not start below 28 and finish at 28 or greater are considered to have experienced reliable change. The rates of clinically significant change and reliable change for adolescents in the full PBH sample were 29% and 45% respectively. For the depression sample of adolescents reporting clinical levels of psychological distress, rates of clinically significant change and reliable change were 56% and 67% respectively. These results are displayed in Table 9.

Table 9

Clinically Significant Change and Reliable Change

	<i>N</i>	Clinically Significant Change	%	Reliable Change	%
PBH full sample					
13-17 years old	2124	612	29	950	45
6-12 years old	2244	713	32	832	37
PBH depression sample					
13-17 years old	270	152	56	181	67
6-12 years old	199	90	45	104	52

Notes: PBH = Public Behavioral Health; CSC = Clinically Significant Change; RC = Reliable Change.

Since a RCI has not been evaluated for the CORS, I calculated a RCI (8.38) based on youth 6 to 12 years old using the current full PBH data set and using formulas from Jacobson & Truax (1991). Combining an estimated and slightly more conservative RCI of 8.4 with the clinical cut-off score of 32 for the CORS, rates of clinically significant change and reliable change for youth 6 to 12 years old are also reported in Table 9. The rates of clinically significant change and reliable change for 6-12 year olds in the full PBH sample were 32% and 37%, respectively. For the depression sample of 6-12 year olds reporting clinical levels of psychological distress, rates of clinically significant change and reliable change were 45% and 52% respectively.

Chapter Four: Discussion

Currently, few researchers have investigated the effectiveness of client feedback with youth involved in mental health treatment in public behavioral health. To my knowledge, this study presents the first benchmarking analysis of treatment outcomes for youth providing client feedback in psychotherapy at a public behavioral health agency. These analyses confirmed three of the four hypotheses. First, I found that the magnitude of change in treatment outcomes from the full PBH sample was (a) clinically equivalent to a treatment group benchmark from all nine client feedback RCTs (OQ System and PCOMS combined, Hypothesis 1a), but not equivalent to the three PCOMS feedback treatment groups alone (Hypothesis 1b); and (b) clinically superior to TAU benchmarks from client feedback RCTs with adult populations (Hypothesis 2a and 2b). Second, I similarly examined the PBH depression sample and found that the magnitude of change in treatment of youth with clinical levels of depression-related psychological distress was (a) clinically superior to a waitlist control benchmark derived from RCT waitlist conditions (Hypothesis 3), and (b) clinically equivalent to the efficacy benchmark based on RCTs of treatments for depression (Hypothesis 4).

In addition, preliminary analyses were conducted on the following client demographic variables: age group (i.e., child or adolescent), race/ethnicity, sex category, and initial diagnoses. Treatment outcomes were significantly different based on diagnoses, age group, and racial/ethnic categories, yet were not significantly different for sex classification of clients (i.e., male or female). Specifically, the magnitude of change was significantly larger for youth less than 13 years of age than for those 13 to 17 years of age, yet they both fell within the moderate to large effect size estimate. Similarly,

while the magnitude of change for treatment with youth identifying as Latino(a)/Hispanic or Native American was large, more moderate magnitudes of change were found for treatment with White/Euro-American and Black/African American youth. After post-hoc analyses, youth identifying as Latino(a)/Hispanic had significantly larger mean pre-post change scores than youth identifying as Black/African American or White/Euro-American. The outcome literature is mixed when looking at the relationship between youth ethnicity and mental health/psychosocial treatment outcome but most often ethnicity is not a significant predictor of treatment outcome (see Huey, Tilley, Jones, & Smith, 2014 and Miranda et al., 2005 for reviews) especially when SES is considered in the analysis (see MTA Cooperative Group, 1999; Swanson et al. 2002) and when treatment models are adapted to be culturally sensitive (Miranda et al., 2005). Moreover, when treatment outcome differences by ethnicity did occur for both youth and adults, results favored ethnic minorities as often as they favored Whites (see Huey et al., 2014 for a review). Since PCOMS is designed to be culturally sensitive (Duncan, 2012) in privileging the client voice, the current finding showing significant outcome differences by ethnicity is surprising, but does not contrast the outcome literature. On a final cautionary note, a substantial number of cases in the sample (34.3%) had missing racial/ethnic information, hence limiting the ability to draw conclusions about the association of ethnicity and treatment outcome for this study.

Effectiveness of Client Feedback with Youth in Poverty

The results of this study suggest that psychotherapy treatment using a client feedback system for youth provided in a particular large PBH setting is effective. The magnitude of change estimated in this study surpassed benchmarks based on client

feedback RCTs with adults and RCTs of mental health treatment with youth specific to depression. Not only was PCOMS effective with youth as a whole, irrespective of diagnosis or level of distress, but for youth presenting with clinical levels of psychological distress related to depression, effect size estimates roughly doubled.

Further, another method of evaluating effectiveness was through determining what percentage of youth experienced reliable change or clinically significant change. Although much room for improvement remains, the current rates of change for the PBH sample may be closer to rates found in managed care settings rather than PBH settings. Very few studies have examined agency-wide mental health care outcomes for youth in usual care settings, and less have analyzed reliable and clinically significant change rates. Warren and colleagues (2010, 2012) perhaps represent the most similar studies from which to draw descriptive comparisons, albeit using a parent/guardian-reported measure. They analyzed data from archives (1997–2008) of a public community mental health system and a large private managed care organization—both having tracked outcomes utilizing the Youth Outcome Questionnaire (Y-OQ; Burlingame et al., 2001) as part of routine clinical services. Descriptively, the current full PBH sample rates of reliable change (29%-32%) and clinically significant change (37%-45%) are similar to, or higher than, rates for the youth in Warren et al.'s managed care settings (28.5%-29% and 29%-30%, respectively). Given the SES disparities in mental healthcare previously discussed, this finding is notable. Also, in comparison to the current PBH sample, Warren et al.'s community mental health center samples had roughly similar rates of reliable change (27.5%-32%) yet much lower rates of clinically significant change (15%-16.8%). One important difference from the current PBH sample of youth at or below the poverty line

is that Warren et al.'s community mental health center samples consisted of youth from not only youth in poverty, but also from average SES backgrounds. These comparisons and contrasts are tentatively made in the absence of more comparable studies; the most notable difference is in the measures utilized—parent/guardian-report (e.g., Y-OQ) versus self-report (e.g., ORS/CORS). Although all of these measures are broad-based (i.e., low specificity), parent/guardian-report outcome measures have been shown in post-hoc analysis as producing a smaller magnitude of change than youth self-report measures for youth dealing with depression (see Weisz et al., 2006).

The current study findings, though limited in their generalizability, contrast with much of the extant literature on mental health treatment outcomes for financially disadvantaged youth. Previous studies (e.g., Weiss et al., 1999; Weiss et al., 2003) have demonstrated that mental health care for youth from economically impoverished backgrounds results in significantly worse treatment outcomes than youth with higher SES. Relatedly, other studies have shown that youth from lower SES backgrounds receive a lower quality of mental healthcare (e.g., less thorough assessment, referrals to other service sectors, patient protection, or parental involvement) than their higher SES peers (Brookman-Frazee et al., 2010; Zima et al., 2005). Furthermore, meta-analyses of psychotherapy interventions with youth from economically impoverished backgrounds have found clinically negligible to small effect size estimates (Farahmand et al., 2012; Farahmand et al., 2011; Lipsey & Wilson, 1998). These results are not surprising considering the accumulation of environmental and psychosocial stressors faced by youth in poverty (Evans, 2004) and the findings that youth from families with low SES were up to three times more likely to have mental health problems than those from families with a

high SES (see Reiss, 2013 for a systematic review). What is surprising, though, is the finding of moderate and large effect size estimates in the current benchmarking study sample of youth in poverty.

I know of only one previous benchmarking study of youth in poverty receiving mental healthcare in a PBH setting (Weersing & Weisz, 2002). As described in detail earlier, Weersing and Weisz found that youth being treated for depression at a community mental health center on average had treatment outcomes equivalent to youth in no-treatment/waitlist control groups in RCTs. Their results contrast sharply with the results of this study. Noteworthy in this contrast is that 10 of the 13 RCTs involved in the creation of their benchmark were also included in the depression-related benchmarks of this study.

Clinical Equivalency to Treatment Benchmarks

The effect size estimates of the current study samples were clinically equivalent to client feedback RCTs overall and to depression-related RCTs. Yet, the effect size estimate for the current full client feedback sample ($N = 4,389$) did not achieve clinical equivalence to the treatment benchmark of the three PCOMS adult studies. One explanation for the lack of clinical equivalence has to do with socioeconomic status. The three adult PCOMS studies (Anker et al., 2009; Reese et al., 2009) included samples that were not exclusively from economically impoverished backgrounds. Two of the three studies involved clients at a university counseling center, and the remaining RCT took place at a couples clinic where no SES information was provided. As detailed earlier, wide agreement exists that SES is a robust predictor of treatment outcome. An alternative explanation comes from the finding that while the therapeutic alliance

significantly predicts treatment outcome for both adults and youth, alliance may be more strongly associated with treatment success in adult psychotherapy (Horvath & Symonds, 1991; Martin et al., 2000).

Supporting Previous PCOMS Research

Despite the lack of equivalence with the three PCOMS adult studies, the current results showing moderate to large magnitudes of change are consistent with previous adult studies evaluating PCOMS in different contexts and treatment modalities, including public behavioral health settings (Reese et al., 2014), telephonic-based counseling (Miller et al., 2006); soldiers in group psychotherapy for substance abuse treatment (Schuman et al., 2015), university students in group psychotherapy (Slone et al., 2015), university students in individual therapy (Reese et al., 2009), and couples therapy (Anker et al., 2009; Reese et al., 2010). Additionally, the current finding of effectiveness is consistent with the one PCOMS effectiveness study of youth in mental health care (Cooper et al., 2013). In their study, youth in Ireland between seven and 11 years old ($N = 288$) were involved in school-based counseling that incorporated PCOMS. They found a mean pre-post effect size estimate (Cohen's d) of 1.49 reflecting improvement in psychological distress related to social, emotional or behavioral difficulties. Although Cooper et al. (2013) did not report the SES of each individual student in the study, on the school level, they reported that 27 of the 28 state schools (non-fee-paying) involved in the study were located in urban areas of “multiple disadvantage” (p. 476).

Study Limitations

Several limitations need consideration when interpreting the findings in the present study. First, a pre-post treatment analysis with only one outcome measure limits

the conclusions. Another limitation of this study pertained to therapist effects (i.e., how treatment outcome differs by therapist). Therapist information in connection to each case was not available for the current dataset. Therapist characteristics, therapist belief in feedback, and the way therapists use feedback have all been shown to be relevant to treatment outcome in feedback studies (e.g., Cooper et al., 2013; de Jong, van Sluis, Nugter, Heiser, & Spinhoven, 2012).

Another limitation of this dataset was its lack of including session by session ORS/CORS scores between initial and final scores. Several studies (Lambert et al., 2003; Sapyta, Riemer, & Bickman, 2005) have shown that client feedback may be most effective for clients who have deteriorating scores on outcome measures. Since the dataset only included first and last session scores, analyses were not possible in evaluating outcome trajectories for not-on-track (NOT) cases.

A further limitation of the current study has to do with measure equivalency. Ideally, when investigating treatment effectiveness through benchmarking, the outcome measures would be the same between the published clinical trials and for the current client population. Psychotherapy outcome measures have been categorized on the basis of two criteria: reactivity and specificity (Lambert, Hatch, Kingston, & Edwards, 1986; Minami et al., 2007). The first criterion, reactivity, is characterized as high or low depending on who completes the measure. For example, if a measure is completed by a clinician, observer, or a family member, it has high reactivity (e.g., Hamilton Rating Scale for Depression [HRSD; Hamilton, 1967]). Measures of high reactivity were left out of the creation of the depression-related benchmarks given that low reactivity measures were available in RCTs. A measure has low reactivity if it contains self-report data (e.g.,

ORS/CORS; Beck Depression Inventory [BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961]). The second criterion, specificity, is characterized as high or low depending on the specificity of symptoms or distress measured; measures focused on specific symptoms or diagnoses (e.g., depression, addiction) are considered to have high specificity (e.g., BDI), and measures of global functioning, broader symptoms, or global distress/well-being have low specificity (e.g., ORS/CORS; Global Severity Index of the Symptom Check List–90—Revised [SCL-90–R; Derogatis, 1983]). The ORS and CORS are considered low reactivity-low specificity (LR-LS) measures.

At present, LR-LS benchmarks for youth depression have not been created. Therefore, in a best effort of equivalence to measure treatment outcome, I only included low reactivity (i.e., self-report) measures in the depression-related efficacy benchmarks. However, the lack of equivalency on the specificity categorization limits the methodological rigor of this study.

Alternatively, a case could be made for this lack of measure equivalency actually making the current study a more conservative comparison. Broad-based measures (i.e., low specificity), similar to the ORS/CORS, often produce smaller effect sizes than high specificity measures (Ford, Hutchings, Bywater, Goodman & Goodman, 2009; Lee, Jones, Goodman, & Heyman, 2005). Minami et al. (2007) empirically tested effect size estimate differences between measures of different specificity and reactivity with adults dealing with depression and found that low reactivity-low specificity measures produced smaller effect size estimates. If further research bears out this interpretation with depression-related measures, the current finding of effect size estimates from broad-based

low-specificity measures surpassing benchmarks made up of high specificity depression measures would be even more impressive.

Implications and Future Recommendations

The current study uniquely contributes to the field by showing how continuous client feedback can improve treatment outcomes for youth in poverty receiving psychotherapy services at a public behavioral health agency. Specific to youth, empirical evidence supports the view that systematically including the child and adolescents' voice in clinical planning (i.e., building the therapeutic alliance), enhances treatment outcomes (Karver et al., 2006; McLeod, 2011; Shirk & Karver). Notably, Duncan et al. (2006) compared ORS/CORS scores to other measures of parent, caregiver, and self-report and concluded: "while youth do rate themselves higher (less distressed) than their caretakers, their views are nonetheless positively correlated with caretakers and are reliable and valid markers of treatment of success" (p. 81).

The current study also has implications related to public policy for mental health treatment in the United States. As previously mentioned, mental health problems make up the top five disabilities affecting children in the United States. (Halfon et al., 2012). Also, depression, the third most common cause of death among adolescents (Arias et al., 2003), is especially costly to individuals, families, and communities. Although millions of children in the United States are treated for mental health problems each year with an estimated 247 billion dollar cost (Centers for Disease Control and Prevention, 2013), the majority of children receiving treatment through publicly funded agencies do not show clinical improvement (Warren et al., 2010). Evidence of PCOMS's effectiveness in

improving treatment outcomes increases confidence to recommend this method for incorporation within public behavioral health systems for youth.

Additional rationale supports the continued implementation and evaluation of PCOMS with youth receiving mental health services. First, PCOMS is currently recognized as an evidence-based intervention for adults 18 years old and older by the Substance Abuse Mental Health Services Administration (SAMHSA) and is listed in the National Registry of Evidence-based Programs and Practices. Second, systematic client feedback is strategic in carrying out what the American Psychological Association considers evidence-based practice (APA Presidential Task Force on Evidence-Based Practice, 2006). Lastly, PCOMS operationalizes key principles of social justice within the psychotherapy relationship (e.g., ongoing self-examination by the therapist, sharing power, and privileging the client voice; Goodman et al., 2004) and a recovery-oriented service model (National Consensus Statement on Mental Health Recovery, 2004).

Duncan (2012) succinctly states:

Routinely requesting, documenting, and responding to client feedback transforms power relations in the immediate therapy encounter by privileging client beliefs and goals over culturally biased and insensitive practices...client feedback protocols undermine inequities built into everyday mental health service delivery by redefining whose voice counts. (p. 99)

Although PCOMS is by no means a panacea for social inequities that structurally exist in the mental health field, it does advance a more socially just paradigm in the counseling room, and when adapted at an agency or policy level, has the potential to influence the transformation of mental health care systems.

Moreover, this study can provide important methodological guidance for evaluating the effectiveness of agency-wide psychotherapy within naturalistic settings.

Threats to internal validity (e.g., regressions to the mean, maturation, demand characteristics) are typically unaccounted for in a pre-post effectiveness study. The current benchmarking strategy addressed these limitations through statistical comparisons to both intent-to-treat treatment groups and waitlist control groups from RCTs. However, benchmarking cannot explain the differences or similarities in treatment outcomes between clinical trial and naturalistic settings.

Given the findings of the current study, future research for client feedback as a quality improvement strategy in PBH agencies serving children and adolescents is warranted. I encourage the replication of this benchmarking study and additional RCTs of client feedback with youth in other PBH settings. As the current study exemplifies, benchmarking methodology allows the bridging of the research-practice gap between RCTs and naturalistic studies. External validity is strengthened beyond effectiveness studies and efficacy studies which typically have smaller homogenous samples.

A second recommendation is for the further psychometric evaluation of brief broad-based treatment outcome benchmarks for future benchmarking studies. Two reasons underline this recommendation. First, several broad-based outcome measures (see Deighton et al., 2014 for an overview) are being developed and implemented as more clinicians and scientists see the benefit of routinely monitoring treatment outcomes in mental health care for youth (Hall et al., 2014). Second, when compared to longer measures, clinicians are more likely to utilize brief measures (Duncan & Reese, 2013)—a key consideration if the field is to further bridge the research-practice gap.

Conclusions

The current study expands our understanding of a client feedback system with ethnically diverse youth receiving mental health services in public behavioral health for a wide range of presenting emotional and behavioral difficulties. The findings of moderate to large effect sizes and overall clinical equivalency with RCT treatment group outcomes support the contention that PCOMS, as a form of systematic client feedback, may be an effective quality improvement strategy in mental health services with youth in public behavioral health—a sector of mental health previously considered to be “in shambles” by the President’s New Freedom Commission on Mental Health (2002, p. ii). Despite the existing socioeconomic disparities in mental healthcare for youth, these findings demonstrate that mental health services to youth in poverty across an entire agency can be effective. Further, results from this study support further expansion of PCOMS in mental health care to address the dire needs of youth who are dealing with clinical levels of psychological distress related to depression.

Appendix A

Outcome Rating Scale (ORS)

Outcome Rating Scale (ORS)

Name _____ Age (Yrs): _____
ID# _____ Sex: M / F
Session # _____ Date: _____

Looking back over the last week, including today, help us understand how you have been doing in the following areas of your life, where marks to the left represent low levels and marks to the right indicate high levels.

Individually:
(Personal well-being)

I-----I

Interpersonally:
(Family, close relationships)

I-----I

Socially:
(Work, School, Friendships)

I-----I

Overall:
(General sense of well-being)

I-----I

Institute for the Study of Therapeutic Change

www.talkingcure.com

© 2000, Scott D. Miller and Barry L. Duncan

Appendix B

Child Outcome Rating Scale (CORS)

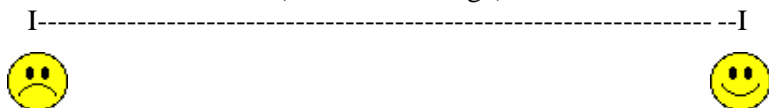
Child Outcome Rating Scale (CORS)

Name _____ Age (Yrs): _____
Sex: M / F _____
Session # _____ Date: _____

How are you doing? How are things going in your life? Please make a mark on the scale to let us know. The closer to the smiley face, the better things are. The closer to the frowny face, things are not so good.

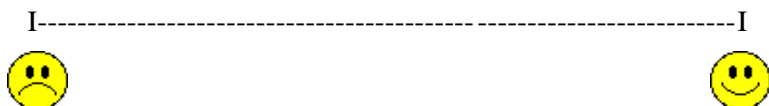
Me

(How am I doing?)



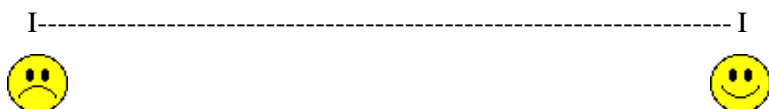
Family

(How are things in my family?)



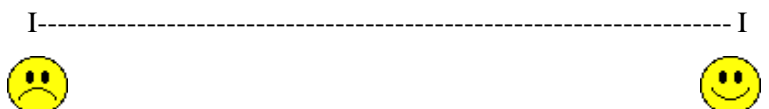
School

(How am I doing at school?)



Everything

(How is everything going?)



Institute for the Study of Therapeutic Change

www.talkingcure.com

© 2003, Barry L. Duncan, Scott D. Miller, & Jacqueline A. Sparks

Appendix C

IRB Review Correspondence

From: Brown, Joe R
To: Jonathan Kodet

Mon, Mar 2, 2015 at 5:42 AM

On February 25, 2015, the Institutional Review Board (IRB) reviewed your project described below.

The IRB determined that your proposal does not meet the Department of Health and Human Services (DHHS) definition of human subjects, a living individual about whom an investigator conducting research obtains (1) data through intervention or interaction with the individual, or (2) identifiable private information” [45 CFR 46.102(f)], and thus does not need IRB review. The IRB made this determination because:

- You will receive existing secondary data;
- You will not receive a link to subjects;
- Subjects cannot be identified based on the variables collected; and
- The data was not collected specifically for your project.

Although your project does not need IRB review, please call the Office of Research Integrity before making any changes to your project because some changes may make the project eligible for IRB review.

If you have any questions regarding the IRB’s decision or if any the information listed above are incorrect, please give the Office of Research Integrity a call at 859-257-9084.

Joe R. Brown, MHS
Research Privacy Specialist
Office of Research Integrity
Office of the Executive Vice President for Research
University of Kentucky
405 Kinhead Hall
Lexington, KY 40506-0057

Phone: (859) 257-9084
Fax: (859) 257-8995

References

- American Psychological Association. (2013). Recognition of psychotherapy effectiveness. *Psychotherapy, 50*, 102-109. doi:10.1037/a0030276
- American Psychological Association Presidential Task Force on Evidence-Based Practice (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271–285. doi:10.1037/0003-066X.61.4.271
- American Psychological Association, Task Force on Socioeconomic Status. (2007). *Report of the APA Task Force on Socioeconomic Status*. Washington, DC: American Psychological Association.
- Ackerman, S. J., Benjamin, L. S., Beutler, L. E., Gelso, C. J., Goldfried, M. R., Hill, C., ... Rainer, J. (2001). Empirically supported therapy relationships: Conclusions and recommendations of the Division 29 Task Force. *Psychotherapy: Theory, Research, Practice, Training, 38*, 495-497. doi:10.1037/0033-3204.38.4.495
- Anker, M. G., Owen, J., Duncan, B. L., & Sparks, J. A. (2010). The alliance in couple therapy: Partner influence, early change, and alliance patterns in a naturalistic sample. *Journal of Consulting and Clinical Psychology, 78*, 635.
doi:[10.1037/a0020051](https://doi.org/10.1037/a0020051)
- Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology, 77*, 693–704.
doi:10.1037/a0016062
- Bickman, L. (1996). A continuum of care: More is not always better. *American Psychologist, 51*, 689–701. doi:10.1037/0003-066X.51.7.689

- Bickman, L., Athay, M. M., Riemer, M., Lambert, E. W., Kelley, S. D., Breda, C., ... Vides de Andrade, A. R. (2010). *Manual of the Peabody treatment progress battery*. Nashville, TN: Vanderbilt University.
- Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 1114-1119. doi:10.1097/CHI.0b013e3181825af8
- Bickman, L., Kelley, S.D., Breda, C., de Andrade, A.R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services*, 62, 1423–1429.
doi:10.1176/appi.ps.002052011
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology* 41, 257–278.
doi:10.1111/j.2044-8317.1988.tb00901.x
- Bodden, D. H. M., Bögels, S. M., Nauta, M. H., De Haan, E., Ringrose, J., Appelboom, C., ... Appelboom-Geerts, C. M. M. J. (2008). Child versus family cognitive-behavioral therapy in clinically anxious youth: An efficacy and partial effectiveness study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47, 1384–1394. doi:10.1097/CHI.0b013e318189148e
- Bohanske, R., & Franczak, M. (2010). Transforming public behavioral health care: A case example of consumer directed services, recovery, and the common factors. In B. Duncan, S. Miller, B. Wampold, & M. Hubble (Eds.), *The heart and soul of change: Delivering what works* (2nd ed., pp. 299–322). doi:10.1037/12075-010

- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice*, 16, 252-260.
doi:10.1037/h0085885
- Borduin, C. M., Mann, B. J., Cone, L. T., Henggeler, S. W., Fucci, B. R., Blaske, D. M., & Williams, R. A. (1995). Multisystemic treatment of serious juvenile offenders: Long-term prevention of criminality and violence. *Journal of Consulting and Clinical Psychology*, 63, 569–578.
- Bosma, H., van de Mheen, H. D., & Mackenbach, J. P. (1999). Social class in childhood and general health in adulthood: Questionnaire study of contribution of psychological attributes. *British Medical Journal*, 318, 18-22.
doi:10.1136/bmj.318.7175.18
- Bringhurst, D. L., Watson, C. W., Miller, S. D., & Duncan, B. L. (2006). The reliability and validity of the Outcome Rating Scale: A replication study of a brief clinical measure. *Journal of Brief Therapy*, 5, 23-30. Retrieved from <http://scottdmiller.com/uploadedFiles/JBTORSReplication.pdf>
- Brookman-Frazee, L., Haine, R. A., Baker-Ericzén, M., Zoffness, R., & Garland, A. F. (2010). Factors associated with use of evidence-based practice strategies in usual care youth psychotherapy. *Administration and Policy in Mental Health*, 37, 254-69. doi:10.1007/s10488-009-0244-9
- Brown, J., Dreis, S., & Nace, D.K. (1999). What really makes a difference in psychotherapy outcome? Why does managed care want to know? In M.A. Hubble, B.L. Duncan, and S.D. Miller (Eds.), *The heart and soul of change: What*

- works in therapy* (pp. 389-406). Washington, DC: American Psychological Association.
- Burlingame, G.M., Mosier, J.I., Wells, M.G., Atkin, Q.G., Lambert, M.J., Whoolery, M., & Latkowski, M. (2001). Tracking the influence of mental health treatment: The development of the Youth Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 8, 361-379. doi:10.1002/cpp.315
- Camp, R. C. (1989). *Benchmarking: The search for industry best practices that lead to superior performance*. Milwaukee, WI: Quality Press.
- Campaign for Mental Health Care Reform. (2005). *Emergency response: A roadmap for federal action on America's mental health crisis*. Retrieved from <http://www.incrisis.org/Documents/EmergencyResponseReport.pdf>
- Campbell, A., & Hemsley, S. (2009). Outcome Rating Scale and Session Rating Scale in psychological practice: Clinical utility of ultra-brief measures. *Clinical Psychologist*, 13, 1-9. doi:10.1080/13284200802676391
- Case, A., & Paxson, C. H. (2006). Children's health and social mobility. *The Future of Children*, 16, 151-173. doi:10.1353/foc.2006.0014
- Centers for Disease Control and Prevention. (2013). Mental Health Surveillance Among Children — United States, 2005–2011. *MMWR. Morbidity and Mortality Weekly Reports*. Retrieved from <http://www.cdc.gov/mmwr/preview/mmwrhtml/su6202a1.htm>
- Clarke, G. N., Hornbrook, M., Lynch, F., Polen, M., Gale, J., Beardslee, W., ... Seeley, J. (2001). A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. *Archives of General*

- Psychiatry*, 58, 1127–1134. Retrieved from
<http://archpsyc.jamanetwork.com/journal.aspx>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, S., Janicki-Deverts, D., Chen, E., & Matthews, K.A. (2010). Childhood socioeconomic status and adult health. *Annals of the New York Academy of Sciences*, 1186(2010), 37-55. doi:10.1111/j.1749-6632.2009.05334.x
- Connors, G. J., Carroll, K. M., DiClemente, C. C., Longabaugh, R., & Donovan, D. M. (1997). The therapeutic alliance and its relationship to alcoholism treatment participation and outcome. *Journal of Consulting and Clinical Psychology*, 65, 588. doi:10.1037//0022-006X.65.4.588
- Cooper, M., Stewart, D., Sparks J., & Bunting, L. (2013). School-based counseling using systematic feedback: A cohort study evaluating outcomes and predictors of change, *Psychotherapy Research*, 23, 474-488, doi:10.1080/10503307.2012.735777
- Curtis, N. M., Ronan, K. R., & Borduin, C. M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology*, 18, 411–419. doi:10.1037/0893-3200.18.3.411
- Curtis, N. M., Ronan, K. R., Heiblum, N., & Crellin, K. (2009). Dissemination and effectiveness of multisystemic treatment in New Zealand: A benchmarking study. *Journal of Family Psychology*, 23, 119-129. doi:10.1037/a0014974
- Curtis, S. E. (1992). *Cognitive-behavioral treatment of adolescent depression*. Unpublished doctoral dissertation, Utah State University, Logan, UT.

- Dana, E. (1998). *A cognitive-behavioral intervention for conduct disordered and concurrently conduct disordered and depressed children*. Unpublished doctoral dissertation, Adelphi University School of Social Work, Garden City, NY.
- Day, C., Carey, M., & Surgenor, T. (2006). Children's key concerns: Piloting a qualitative approach to understanding their experience of mental health care. *Clinical Child Psychology and Psychiatry*, *11*, 139-155.
doi:10.1177/1359104506056322
- De Cuyper, S., Timbremont, B., Braet, C., De Backer, V., & Wullaert, T. (2004). Treating depressive symptoms in school children: A pilot study. *Journal of European Child and Adolescent Psychiatry*, *13*, 105–114.
- Diamond, G. S., Reis, B. F., Diamond, G. M., Siqueland, L., & Isaacs, L. (2002). Attachment-based family therapy for depressed adolescents: A treatment development study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*, 1190–1196.
- Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: a review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health*, *8*, 1-14. Retrieved from <http://www.capmh.com/content/8/1/14>
- Dohrenwend, B. P., Levav, I., Shrout, P. E., Schwartz, S., Naveh, G., Link, B. G., ... Stueve, A. (1992). Socioeconomic status and psychiatric disorders: the causation-selection issue. *Science*, *255*, 946-952. doi:10.1126/science.1546291

- Dohrenwend, B. P. (1990). Socioeconomic status (SES) and psychiatric disorders. *Social Psychiatry and Psychiatric Epidemiology*, 25, 41-47.
doi:10.2105/AJPH.2011.300477
- Duncan, B. (2011). *The partners for change outcome management system (PCOMS): Administration, scoring, interpreting update for the Outcome and Session Ratings Scale*. Jensen Beach, FL: Author.
- Duncan, B. (2012). The partners for change outcome management system (PCOMS): The heart and soul of change project. *Canadian Psychology*, 53, 93-104.
doi:10.1037/a0027762
- Duncan, B. L. (2014). *On becoming a better therapist: Evidence-based practice one client at a time*, (2nd ed.). Baltimore, MD: American Psychological Association.
- Duncan, B. L., Miller, S. D., & Sparks, J. A. (2004). *The heroic client: A revolutionary way to improve effectiveness through client-directed outcome-informed therapy* (rev. ed.). San Francisco, CA: Jossey-Bass.
- Duncan, B., Miller, S., Sparks, J., Claud, D., Reynolds, L., Brown, J., & Johnson, L. (2003). The Session Rating Scale: Preliminary psychometric properties of a “working” alliance measure. *Journal of Brief Therapy*, 3, 3–12. Retrieved from <https://heartandsoulofchange.com/content/resources/viewer.php?resource=article&id=23>
- Duncan, B., Miller, S., Wampold, B., & Hubble, M. (Eds.). (2010). *The heart and soul of change: Delivering what works in therapy* (2nd ed.). Washington, DC: American Psychological Association. doi:10.1037/12075-000

- Duncan, B., Sparks, J., Miller, S., Bohanske, R., & Claud, D. (2006). Giving youth a voice: A preliminary study of the reliability and validity of a brief outcome measure for children. *Journal of Brief Therapy*, 5, 71-87. Retrieved from <https://heartandsoulofchange.com/content/resources/viewer.php?resource=article&id=24>
- Edwards, A. N. (2014). *Dynamics of economic well-being: Poverty, 2009–2011*. (Report No. P70-137). Retrieved from The U.S. Census Bureau website: <https://www.census.gov/library/publications/2014/demo/p70-137.html>
- Ehrenreich-May, J., Southam-Gerow, M. A., Hourigan, S. E., Wright, L. R., Pincus, D. B., & Weisz, J. R. (2011). Characteristics of anxious and depressed youth seen in two different clinical contexts. *Administration and Policy in Mental Health and Mental Health Services Research*, 38, 398-411.
- Ettelson, R. G. (2003). *The treatment of adolescent depression*. Unpublished doctoral dissertation, Illinois State University, Normal, IL.
- Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist*, 59, 77-92. doi:10.1037/0003-066X.59.2.77-92.
- Farahmand, F. K., Duffy, S. N., Taylor, M. A., DuBois, D. L., Lyon, A. L., Grant, K. E., ... Nathanson, A. M. (2012). Community-based mental health and behavioral programs for low-income urban youth: A meta-analytic review. *Clinical Psychology: Science and Practice*, 19, 195-215. doi:10.1111/j.1468-2850.2012.01283.x
- Farahmand, F. K., Grant, K. E., Polo, A. J., Duffy, S. N., & DuBois, D. L. (2011). School-based mental health and behavioral programs for low-income, urban

- youth: A systematic and meta-analytic review. *Clinical Psychology: Science & Practice*, 18, 372–390. doi:10.1111/j.1468-2850.2011.01265.x
- Farmer, E. Z., & Burns, B. J. (1997). Impact of children's mental health problems on families. *Journal of Emotional & Behavioral Disorders*, 5, 230–238. doi: 10.1177/106342669700500406
- Farrell, L. J., Schlup, B., & Boschen, M. (2010). Cognitive-behavioral treatment of childhood obsessive-compulsive disorder in community-based clinical practice: Clinical significance and benchmarking against efficacy. *Behaviour Research and Therapy*, 48, 409–417. doi:10.1016/j.brat.2010.01.004
- Fischer, S. A. (1995). *Development and evaluation of group cognitive-behavioral therapy for depressed and suicidal adolescents in juvenile detention*. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.
- Ford, T., Hutchings, J., Bywater, T., Goodman A., & Goodman, R. (2009). Evaluating effectiveness in child mental health interventions. *British Journal of Psychiatry*, 194, 552-558. doi: 10.1192/bjp.bp.108.052373
- Frank, R. G., & Glied, S. A. (2006). *Better but not well: Mental health policy in the United States since 1950*. Baltimore: MD: Johns Hopkins University Press.
- Fryers, T., Melzer, D., & Jenkins, R. (2003). Social inequalities and the common mental disorders. *Social Psychiatry and Psychiatric Epidemiology*, 38, 229-237. doi:10.1007/s00127-003-0627-2
- Garcia, J. A., & Weisz, J. R. (2002). When youth mental health care stops: Therapeutic relationship problems and other reasons for ending youth outpatient treatment.

Journal of Consulting and Clinical Psychology, 70, 439. doi:10.1037/0022-006X.70.2.439

Gallo, L. C., Bogart, L. M., Vranceanu, A. M., & Matthews, K. A. (2005).

Socioeconomic status, resources, psychological experiences, and emotional responses: A test of the reserve capacity model. *Journal of Personality and Social Psychology*, 88, 386. doi:10.1037/0022-3514.88.2.386

Garland, A. F., Accurso, E. C., Haine-Schlagel, R., Brookman-Frazee, L., Roesch, S., &

Zhang, J. J. (2014). Searching for elements of evidence-based practices in children's usual care and examining their impact. *Journal of Clinical Child & Adolescent Psychology*, 43, 201-215. doi:10.1080/15374416.2013.869750

Garland, A. F., Haine-Schlagel, R., Brookman-Frazee, L., Baker-Ericzen, M., Trask, E.,

& Fawley-King, K. (2013). Improving community-based mental health care for children: Translating knowledge into action. *Administration and Policy in Mental Health and Mental Health Services Research*, 40, 6-22. doi:10.1007/s10488-012-0450-8

Gillaspy, J. A., & Murphy, J. J. (2011). The use of ultra-brief client feedback tools in

SFBT. In C. W. Franklin, T. Trepper, E. McCollum, & W. Gingerich (Eds.). *Solution-focused brief therapy* (pp 73-94). New York, NY: Oxford University Press.

Goodman, J. D., McKay, J. R., & DePhilippis, D. (2013). Progress monitoring in mental

health and addiction treatment: A means of improving care. *Professional Psychology: Research and Practice*, 44, 231-246. doi:10.1037/a0032605

- Gould, M. S., Shaffer, D., & Kaplan, D. (1985). The characteristics of dropouts from a child psychiatry clinic. *Journal of the American Academy of Child Psychiatry*, 24, 316–328. doi:10.1016/S0002-7138(09)61093-1
- Halfon, N., Houtrow, A., Larson, K., & Newacheck, P. W. (2012). The changing landscape of disability in childhood. *The Future of Children*, 22, 13-42. doi:10.1353/foc.2012.0004
- Harpaz-Rotem, I., Leslie, D., & Rosenheck, R. A. (2004). Treatment and retention among children entering a new episode of mental health care. *Psychiatric Services*, 55, 1022–1028. doi:10.1176/appi.ps.55.9.1022
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Henggeler, S. W., Melton, G. B., Brondino, M. J., Scherer, D. G., & Hanley, J. H. (1997). Multisystemic therapy with violent and chronic juvenile offenders and their families: The role of treatment fidelity in successful dissemination. *Journal of Consulting and Clinical Psychology*, 65, 821–833.
- Henggeler, S. W., Melton, G. B., & Smith, L. A. (1992). Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting and Clinical Psychology*, 60, 953–961.
- Hill, C.E., & Lambert, M.J. (2004). Methodological issues in studying psychotherapy processes and outcomes. In M.J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 84-135). Chicago, IL: John Wiley and Sons.

- Horvath, A. O. & Symonds, D. B. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology*, 38, 139–149. doi:10.1037/0022-0167.38.2.139
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, 48, 9-16. doi:10.1037/a0022186
- Hudson, C. G. (2005). Socioeconomic status and mental illness: Tests of the social causation and selection hypotheses. *American Journal of Orthopsychiatry*, 75, 3–18. doi:10.1037/0002-9432.75.1.3
- Huey Jr, S. J., Tilley, J. L., Jones, E. O., & Smith, C. A. (2014). The contribution of cultural competence to evidence-based care for ethnically diverse populations. *Annual Review of Clinical Psychology*, 10, 305-338. doi:10.1146/annurev-clinpsy-032813-153729
- In-Albon, T., & Schneider, S. (2007). Psychotherapy of child anxiety disorders: A meta-analysis. *Psychotherapy and Psychosomatics*, 76, 15–24. doi:10.1159/000096361
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19. doi:10.1037/0022-006X.59.1.12
- Jokela, M., Batty, G. D., Vahtera, J., Elovainio, M., & Kivimäki, M. (2013). Socioeconomic inequalities in common mental disorders and psychotherapy treatment in the UK between 1991 and 2009. *The British Journal of Psychiatry*, 202, 115–120. doi:10.1192/bjp.bp.111.098863

- Kahn, J. S., Kehle, T. J., Jenson, W. R., & Clark, E. (1990). Comparison of cognitive-behavioral, relaxation, and self-modeling interventions for depression among middle-school students. *School Psychology Review* 19, 196–211.
- Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2006). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: Rand Corporation.
- Karver, M. S., Handelsman, J. B., Fields, S., & Bickman, L. (2006). Meta-analysis of therapeutic relationship variables in youth and family therapy: The evidence for different relationship variables in the child and adolescent treatment outcome literature. *Clinical Psychology Review*, 26, 50–65. 10.1016/j.cpr.2005.09.001
- Kessler, R. C., & Cleary, P. D. (1980). Social class and psychological distress. *American Sociological Review*, 45, 463-478. doi:10.2307/2095178
- Klein, J. B., Jacobs, R. H., & Reinecke, M. A. (2007). Cognitive-behavioral therapy for adolescent depression: A meta-analytic investigation of changes in effect-size estimates. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46, 1403-1413. doi:10.1097/chi.0b013e3180592aaa
- Lambert, M. (2007) Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17, 1–14. doi:10.1080/10503300601032506
- Lambert, M. J. (2010). “Yes, it is time for clinicians to monitor treatment outcome”. In B. L. Duncan, S. C. Miller, B. E. Wampold, & M. A. Hubble (Eds.), *Heart and soul of change: Delivering what works in therapy* (2nd ed., pp. 239–266). Washington, DC: American Psychological Association.

- Lambert, M. J. (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed., pp. 169–218). Hoboken, NJ: Wiley.
- Lambert, W., & Bickman, L. (2004). The Clock-Setting Cure: How children's symptoms might improve after ineffective treatment. *Psychiatric Services*, 55, 381-382.
doi:10.1176/appi.ps.55.4.381
- Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G. M., & Reisinger, C. (1996). *Administration and scoring manual for the OQ 45.2*. Stevenson, MD: American Professional Credentialing Services.
- Lambert, M. J., Hatch, D. R., Kingston, M. D., & Edwards, B. C. (1986). Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: A meta-analytic comparison. *Journal of Consulting and Clinical Psychology*, 54, 54-59.
doi:10.1037//0022-006X.54.1.54
- Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, 48, 72-79. doi:10.1037/a0022238
- Lambert, M. J. & Ogles, B.M. (2009) Using clinical significance in psychotherapy outcome research: The need for a common procedure and validity data. *Psychotherapy Research*, 19, 493-501, doi:10.1080/10503300902849483
- Laska, K. M., Gurman, A. S., & Wampold, B. E. (2014). Expanding the lens of evidence-based practice in psychotherapy: A common factors perspective. *Psychotherapy*, 51, 467-481. doi:10.1037/a0034332

- Lau, W. Y., Chan, C. K., Li, J. C., & Au, T. K. (2010). Effectiveness of group cognitive-behavioral treatment for childhood anxiety in community clinics. *Behaviour Research and Therapy*, 48, 1067–1077. doi:10.1016/j.brat.2010.07.007
- Law, D., & Wolpert, M. (Eds.) (2014). *Guide to using outcomes and feedback tools with children, young people and families* (2nd ed). London, England: CAMHS Press.
- Lee, C. M., Horvath, C., & Hunsley, J. (2013). Does it work in the real world? The effectiveness of treatments for psychological problems in children and adolescents. *Professional Psychology: Research and Practice*, 44, 81. doi:10.1037/a0031133
- Lee, W., Jones, L., Goodman, R., & Heyman, I. (2005). Broad outcome measures may underestimate effectiveness: an instrument comparison study. *Child and Adolescent Mental Health*, 10, 143-144. doi:10.1111/j.1475-3588.2005.00350.x
- Lewinsohn, P. M., & Clarke, G. N. (1999). Psychosocial treatments for adolescent depression. *Clinical Psychology Review*, 19, 329–342. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0272735898000555>
- Lewinsohn, P. M., Clarke, G. N., Hops, H., & Andrews, J. (1990). Cognitive-behavioral treatment for depressed adolescents. *Behavior Therapy*, 21, 385–401. doi:10.1016/S0005-7894(05)80353-3
- Liddle, B., & Spence, S. H. (1990). Cognitive-Behaviour Therapy with depressed primary school children: A cautionary note. *Behavioural Psychotherapy*, 18, 85-102. doi: 10.1017/S0141347300018218
- Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington

- (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 313-345). Thousand Oaks, CA: Sage Publications.
- Lorant, V., Delière, D., Eaton, W., Robert, A., Philpott, P., & Ansseau, M. (2003). Socioeconomic inequalities in depression: A meta-analysis. *American Journal of Epidemiology*, 157, 98-112. doi:10.1093/aje/kwf182
- Lundberg, O. (1997). Childhood conditions, sense of coherence, social class and adult ill health: Exploring their theoretical and empirical relations,” *Social Science and Medicine*, 44, 821-831. doi:10.1016/S0277-9536(96)00184-0
- Manteuffel, B., Stephens, R. L., Sondheimer, D. L., & Fisher, S. K. (2008). Characteristics, service experiences, and outcomes of transition-aged youth in systems of care: Programmatic and policy implications. *The Journal of Behavioral Health Services Research*. 35, 469–487. doi:10.1007/s11414-008-9130-6
- Marcotte, D., & Baron, P. (1993). L’efficacite d’une strategie d’intervention emotivo-rationnelle aupres d’adolescents depressifs du milieu scolaire [The efficacy of a school-based rational-emotive intervention strategy with depressive adolescents]. *Canadian Journal of Counselling*, 27, 77–92. Retrieved from <http://cjc-rcc.ucalgary.ca/cjc/index.php/rcc/article/view/545>
- Martin, D. J., Garske, F. P., & Davis, M. K. (2000). Relationship of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68, 438–450. doi:10.1037//0022-006X.68.3.438

- McHugh, K. R., & Barlow, D. H. (2012). *Dissemination and implementation of evidence-based psychological interventions*. New York, NY:Oxford University Press.
- McLaughlin, K. A., Costello, E. J., Leblanc, W., Sampson, N. A., & Kessler, R. C. (2012). Socioeconomic status and adolescent mental disorders. *American Journal of Public Health, 102*, 1742-1750. doi:10.2105/AJPH.2011.300477
- McLeod, B. D. (2011). Relation of the alliance with outcomes in youth psychotherapy: A meta-analysis. *Clinical Psychology Review, 31*, 603–616.
doi:10.1016/j.cpr.2011.02.001
- McLeod, J. D., & Shanahan, M. J. (1993). Poverty, parenting and children's mental health. *American Sociological Review, 58*, 1993, 351-366. doi:10.2307/2095178
- Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., ... Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry, 49*, 980-989. doi:10.1016/j.jaac.2010.05.017
- Merrill, K. A., Tolbert, V. E., & Wade, W. A. (2003). Effectiveness of cognitive therapy for depression in a community mental health center: A benchmarking study. *Journal of Consulting and Clinical Psychology, 71*, 404-409. doi:10.1037/0022-006X.71.2.404
- Michael, K. D., & Crowley, S. L. (2002). How effective are treatments for child and adolescent depression? A meta-analytic review. *Clinical Psychology Review, 22*, 247–269. Retrieved from
<http://www.sciencedirect.com/science/journal/02727358/22/2>

- Miller, S. D., & Duncan, B. L. (2004). *The Outcome and Session Rating Scales: Administration and scoring manual*. Jensen Beach, FL: Authors.
- Miller, S. D., Duncan, B. L., Brown, J., Sorrell, R., & Chalk, B. (2006). Using formal client feedback to improve retention and outcome: Making ongoing, real-time assessment feasible. *Journal of Brief Therapy*, 5, 5-22. Retrieved from <http://journalbrieftherapy.com>
- Miller, S. D., Mee-Lee, D., Plum, W., & Hubble, M. A. (2005). Making treatment count: Client-directed, outcome informed clinical work with problem drinkers. In J. L. Lebow (Ed.), *Handbook of clinical family therapy* (pp. 281–308). Hoboken, NJ: Wiley.
- Miller, S., Duncan, B., Brown, J., Sparks, J., & Claud, D. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy*, 2, 91-100. Retrieved from <https://heartandsoulofchange.com/content/resources/viewer.php?resource=article&id=26>
- Minami, T., Davies, D. R., Tierney, S. C., Bettmann, J. E., McAward, S. M., Averill, L. A., ... Wampold, B. E. (2009). Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center. *Journal of Counseling Psychology*, 56, 309–320. doi:10.1037/a0015398
- Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. S. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality & Quantity*, 42, 513–525. doi:10.1007/s11135-006-9057-z

- Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology, 76*, 116–124. doi:10.1037/0022-006X.76.1.116
- Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. (2007). Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology, 75*, 232–243. doi:10.1037/0022-006X.75.2.232
- Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. S. (2008). Using clinical trials to benchmark effects produced in clinical practice. *Quality & Quantity, 42*, 513–525. doi:10.1007/s11135-006-9057-z
- Miranda, J., Bernal, G., Lau, A., Kohn, L., Hwang, W. C., & LaFromboise, T. (2005). State of the science on psychosocial interventions for ethnic minorities. *Annual Review of Clinical Psychology, 1*, 113-142. doi:0.1146/annurev.clinpsy.1.102803.143822
- MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 56*, 1073-1086. doi:10.1001/archpsyc.56.12.1073
- Mufson, L., Dorta, K. P., Wickramaratne, P., Nomura, Y., Olfson, M., & Weissman, M. M. (2004). A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry, 61*, 577–584. Retrieved from <http://archpsyc.jamanetwork.com/article.aspx?articleid=482013>

- Mufson, L., Weissman, M. M., Moreau, D., & Garfinkel, R. (1999). Efficacy of interpersonal psychotherapy for depressed adolescents. *Archives of General Psychiatry*, 56, 573–579. doi:10.1001/archpsyc.56.6.573
- Muntaner, C., Eaton, W. W., Miech, R., & O’Campo, P. (2004). Socioeconomic position and major mental disorders. *Epidemiologic Reviews*, 26, 53–62. doi:10.1093/epirev/mxh001
- Nelson, P. L., Warren, J. S., Gleave, R. L., & Burlingame, G. M. (2013). Youth psychotherapy change trajectories and early warning system accuracy in a managed care setting. *Journal of Clinical Psychology*, 69, 880-895. doi:10.1002/jclp.21963
- Orlinsky, D. E., Rønnestad, M. H., & Willutzki, U. (2004). Fifty years of process-outcome research: Continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield’s handbook of psychotherapy and behavior change* (5th ed., pp. 307-390). New York, NY: Wiley.
- Overington, L., & Ionita, G. (2012). Progress monitoring measures: A brief guide. *Canadian Psychology/Psychologie Canadienne*, 53, 82-92. doi:10.1037/a0028017
- Pan, Y. J., Stewart, R., & Chang, C. K. (2013). Socioeconomic disadvantage, mental disorders and risk of 12-month suicide ideation and attempt in the National Comorbidity Survey Replication (NCS-R) in US. *Social Psychiatry and Psychiatric Epidemiology*, 48, 71-79. doi:10.1007/s00127-012-0591-9
- Pfuntner, A., Wier, L. M., & Stocks, C. (2013 January). *Most Frequent Conditions in U.S. Hospitals, 2010*. HCUP Statistical Brief #148. Agency for Healthcare

- Research and Quality, Rockville, MD. Retrieved from <http://www.hcup-us.ahrq.gov/reports/statbriefs/sb148.pdf>
- Pinquart, M., & Sörensen, S. (2000). Influences of socioeconomic status, social network, and competence on subjective well-being in later life: A meta-analysis. *Psychology and Aging, 15*, 187-224. doi:10.1037/0882-7974.15.2.187
- President's New Freedom Commission on Mental Health. (2002). *Interim report* (DHHS Pub. No. SMA-03-3932). Retrieved from http://www.mentalhealthcommission.gov/reports/interim_toc.htm
- Reinecke, M. A., Ryan, N. E., & DuBois, D. L. (1998). Cognitive-behavioral therapy of depression and depressive symptoms during adolescence: A review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*, 26-34. doi: 10.1097/00004583-199801000-00013
- Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Social Science & Medicine, 90*, 24-31. doi:10.1016/j.socscimed.2013.04.026
- Reese, R. J., Duncan, B. L., Bohanske, R. T., Owen, J. J., & Minami, T. (2014). Benchmarking outcomes in a public behavioral health setting: Feedback as a quality improvement strategy. *Journal of Consulting and Clinical Psychology, 82*, 731-742. doi: 10.1037/a0036915
- Reese, R. J., Norsworthy, L. A., & Rowlands, S. R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy: Theory, Research, Practice, Training, 46*, 418-431. doi:10.1037/a0017901

- Reese, R. J., Toland, M. D., Slone, N. C., & Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy: Theory, Research, Practice, Training*, 47, 616–630. doi:10.1037/a0021182
- Reynolds, W. M., & Coats, K. I. (1986). A comparison of cognitive–behavioral therapy and relaxation training for the treatment of depression in adolescents. *Journal of Consulting and Clinical Psychology*, 54, 653–660. doi:10.1037/0022-006X.54.5.653
- Ritsher, J. E., Warner, V., Johnson, J. G., & Dohrenwend, B. P. (2001). Inter-generational longitudinal study of social class and depression: a test of social causation and social selection models. *The British Journal of Psychiatry*, 178, s84-s90. doi:10.1192/bjp.178.40.s84
- Rohde, P., Clarke, G. N., Mace, D. E., Jorgensen, J. S., & Seeley, J. R. (2004). An efficacy/effectiveness study of cognitive– behavioral treatment for adolescents with and without comorbid major depression and conduct disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43, 660–668. doi:10.1097/01.chi.0000121067.29744.41
- Rosello, J., & Bernal, G. (1999). The efficacy of cognitive–behavioral and interpersonal treatments for depression in Puerto Rican adolescents. *Journal of Consulting and Clinical Psychology*, 67, 734–745. doi:10.1037/0022-006X.67.5.734
- Sapyta, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research, and practice. *Journal of Clinical Psychology*, 61, 145-153. doi:10.1002/jclp.20107
- Schuman, D. L., Slone, N. C., Reese, R. J., & Duncan, B. (2015). Efficacy of client feedback in group psychotherapy with soldiers referred for substance abuse

- treatment. *Psychotherapy Research*, 25, 396-407.
doi:10.1080/10503307.2014.900875
- Serlin, R. C., & Lapsley, D. K. (1985). Rationality in psychological research: The good-enough principle. *American Psychologist*, 40, 73-83. doi:10.1037/0003-066X.40.1.73
- Serlin, R. C., & Lapsley, D. K. (1993). Rational appraisal of psychological research and the good-enough principle. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 199–228). Hillsdale, NJ: Erlbaum.
- Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 71, 452-464. doi:10.1037/0022-006X.71.3.452
- Shonkoff, J. P., Boyce, W. T., & McEwen, B. S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities: Building a new framework for health promotion and disease prevention. *Journal of the American Medical Association*, 301, 2252-2259. doi:10.1001/jama.2009.754
- Slone, N. C., Reese, R. J., Mathews-Duvall, S., & Kodet, J. (2015). Evaluating the efficacy of client feedback in group psychotherapy. *Group Dynamics: Theory, Research, and Practice*. 19, 122-136. doi:10.1037/gdn0000026
- Soni, A. (2009). *The five most costly children's conditions, 2006: Estimates for the US civilian noninstitutionalized children, ages 0–17*. Statistical Brief #242. Rockville,

- MD: Agency for Healthcare Research and Quality. Retrieved from
http://www.meps.ahrq.gov/mepsweb/data_files/publications/st242/stat242.pdf.
- Southam-Gerow, M. A., Chorpita, B. F., Miller, L. M., & Gleacher, A. A. (2008). Private and public referrals in a child anxiety disorders clinic: Similarities and differences. *Administration and Policy in Mental Health and Mental Health Services Research*, 35, 168–180.
- Stark, K. D., Reynolds, W. M., & Kaslow, N. J. (1987). A comparison of the relative efficacy of self-control therapy and a behavioral problem-solving therapy for depression in children. *Journal of Abnormal Child Psychology*, 15, 91–113.
doi:10.1007/BF00916468
- Stith, S. M., Rosen, K. H., McCollum, E. E., Coleman, J. U., & Herman, S. A. (1996). The voices of children: Preadolescent children's experiences in family therapy. *Journal of Marital and Family Therapy*, 22, 69-86. doi:10.1111/j.1752-0606.1996.tb00188.x
- Strohschein, L. (2005). Household income histories and child mental health trajectories. *Journal of Health and Social Behavior*, 46, 359-375.
doi:10.1177/002214650504600404
- Timimi, S., Tetley, D., Burgoine, W., & Walker, G. (2013). Outcome orientated child and adolescent mental health services (OO-CAMHS): A whole service model. *Clinical Child Psychology and Psychiatry*, 18, 169-184.
doi:10.1177/1359104512444118

- Tracey, T. J. & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment*, 1, 207–210. doi:10.1037/1040-3590.1.3.207
- Treatment for Adolescents with Depression Study (TADS) Team. (2004). Fluoxetine, cognitive– behavioral therapy, and their combination for adolescents with depression. *Journal of the American Medical Association*, 292, 807–820. doi:10.1001/jama.292.7.807
- Valderhaug, R., Larsson, B., Götestam, G., & Piancentini, J. (2007). An open clinical trial of cognitive-behaviour therapy in children and adolescents with obsessive compulsive disorder administered in regular outpatient clinics. *Behaviour Research and Therapy*, 45, 577–589. doi:10.1016/j.brat.2006.04.011
- Vitiello, B., Rohde, P., Silva, S., Wells, K., Casat, C., Waslick, B., ... March, J. (2006). Functioning and quality of life in the Treatment for Adolescents with Depression Study (TADS). *Journal of the American Academy of Child & Adolescent Psychiatry*, 45, 1419-1426. doi:10.1097/01.chi.0000242229.52646.6e
- Vostanis, P., Feehan, C., Grattan, E., & Bickerton, W. L. (1996). Treatment for children and adolescents with depression: Lessons from a controlled trial. *Clinical Child Psychology and Psychiatry*, 1, 199–212. doi:10.1177/1359104596012003
- Walter, H.J., Gouze, K., Cicchetti, C., Arend, R., Mehta, T., Schmidt, J., & Skvarla, M. (2011). A pilot demonstration of comprehensive mental health services in inner-city public schools. *Journal of School Health*, 81, 185-193. doi:10.1111/j.1746-1561.2010.00578.x

- Wampold, B. E. (2001). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.
- Warren, J. S., Nelson, P. L., Burlingame, G. M., & Mondragon, S. A. (2012). Predicting patient deterioration in youth mental health services: Community mental health vs. managed care settings. *Journal of Clinical Psychology, 68*, 24-40. doi:10.1002/jclp.20831
- Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of Consulting and Clinical Psychology, 78*, 144-155. doi:10.1037/a0018544
- Watanabe, N., Hunot, V., Omori, I. M., Churchill, R., & Furukawa, T. A. (2007). Psychotherapy for depression among children and adolescents: A systematic review. *Acta Psychiatrica Scandinavica, 116*, 84-95. doi:10.1111/j.1600-0447.2007.01018.x
- Weersing, V. R., & Weisz, J. R. (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology, 70*, 299–310. doi:10.1037/0022-006X.70.2.299
- Weiss, B., Catron, T., Harris, V., & Phung, T. M. (1999). The effectiveness of traditional child psychotherapy. *Journal of Consulting and Clinical Psychology, 67*, 82-94. doi:10.1037//0022-006X.67.1.82
- Weiss, B., Harris, V., Catron, T., & Han, S. S. (2003). Efficacy of the RECAP intervention program for children with concurrent internalizing and externalizing

- problems. *Journal of Consulting and Clinical Psychology*, 71, 364-374.
doi:10.1037/0022-006X.71.2.364
- Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*. New York: Cambridge University Press.
- Weisz, J. R. (2014). Building Robust Psychotherapies for Children and Adolescents. *Perspectives on Psychological Science*, 9, 81-84. doi:10.1177/1745691613512658
- Weisz, J. R., & Jensen, A. L. (2001). Child and adolescent psychotherapy in research and practice contexts: Review of the evidence and suggestions for improving the field. *European Child & Adolescent Psychiatry*, 10, s12-s18.
doi:10.1007/s007870170003
- Weisz, J. R., Jensen Doss, A., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*, 56, 337–363. doi:10.1146/annurev.psych.55.090902.141449
- Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: a meta-analysis. *Psychological Bulletin*, 132, 132-149. doi:10.1037/0033-2909.132.1.132-149
- Weisz, J. R., Southam-Gerow, M. A., Gordis, E. B., Connor-Smith, J. K., Chu, B. C., Langer, D. A., ... Weiss, B. (2009). Cognitive-behavioral therapy versus usual clinical care for youth depression: An initial test of the transportability to community clinics and clinicians *Journal of Consulting and Clinical Psychology*, 77, 383–396. doi:10.1037/a0013877
- Weisz, J. R., Thurber, C. A., Sweeney, L., Proffitt, V. D., & LeGagnoux, G. L. (1997). Brief treatment of mild-to-moderate child depression using primary and

- secondary control enhancement training. *Journal of Consulting and Clinical Psychology*, 65, 703–707. doi:10.1037/0022-006X.65.4.703
- Weisz, J. R., Weiss, B., & Langmeyer, D. B. (1987). Giving up on child psychotherapy: Who drops out? *Journal of Consulting and Clinical Psychology*, 55, 916–918. doi:10.1037//0022-006X.55.6.916
- Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identification of treatment failure and problem-solving strategies in routine practice. *Journal of Counseling Psychology*, 50, 59-68. doi:10.1037/0022-0167.50.1.59
- Wierzbicki, M., & Pekarik, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice*, 24, 190-195. doi:10.1037//0735-7028.24.2.190
- Wolpert, M., Cheng, H., & Deighton, J. (2014). Measurement Issues: Review of four patient reported outcome measures: SDQ, RCADS, C/ORS and GBO—their strengths and limitations for clinical use and service evaluation. *Child and Adolescent Mental Health*. Advance online publication. doi:10.1111/camh.12065
- World Health Organization. (2012). Maternal and child mental health. Retrieved from http://www.who.int/mental_health/maternal-child/en/
- Zima, B. T., Hurlburt, M. S., Knapp, P., Ladd, H., Tang, L., Duan, N., ... Wells, K. B. (2005). Quality of publicly-funded outpatient specialty mental health care for common childhood psychiatric disorders in California. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44, 130-144. doi:10.1097/00004583-200502000-00005

- Zima, B. T., & Mangione-Smith, R. (2011). Gaps in quality measures for child mental health care: An opportunity for a collaborative agenda. *Journal of the American Academy of Child & Adolescent Psychiatry*, 50, 735-737.
doi:10.1016/j.jaac.2011.05.006
- Zirkelback, E. A., & Reese, R. J. (2010). A review of psychotherapy outcome research: Considerations for school-based mental health providers. *Psychology in the Schools*, 47, 1084-1100. doi:10.1002/pits.20526

Curriculum Vitae

Jonathan Kodet, MS, EdS, AFC®

EDUCATION

- 2011-2013 **Education Specialist (EdS) in Counseling Psychology**
University of Kentucky; Lexington, KY
- 2004-2006 **Masters of Science in Human Services, Specialization in Counseling Studies**
Capella University; Minneapolis, MN
Thesis Title: *Juvenile Delinquency Recidivism and Group Rational Emotive Behavior Therapy*
- 1997-2000 **Bachelor of Arts in Theology and Youth Ministries**
Bethel University; St. Paul, MN
Paper Titles: 1) *Mutually Vulnerable Interreligious Dialogue*, 2) *A Theology of Liberation: Gustavo Gutierrez*

SUPERVISED CLINICAL AND SUPERVISION EXPERIENCE

- August 2015-present **University of San Diego Counseling Center (APA-accredited)**
Full-time Predoctoral Intern
Site Supervisor/Coordinator of Training: Karen Lese-Fowler, PhD
- August 2014-December 2014 **University of Kentucky Counseling Center**
Therapist and Group Co-Leader, Post-Masters Practicum
Site Supervisors: Nathan Miles, PhD, Tina Bryant, PhD
Faculty Supervisor: Jeff Reese, PhD
- Sept 2014-December 2014 **University of Kentucky Counseling Psychology Program, Lexington, KY**
Supervisor of Masters Students, Post-Masters Supervision Practicum
Faculty Supervisors: Rory Remer PhD
- August 2013-May 2014 **Bluegrass.org, Newtown Counseling Center (Community Mental Health Center), Lexington, KY**
Psychology Practicum Student, Post-Masters Practicum
Site Supervisor: Pat Burke, PhD
Faculty Supervisor: Keisha Love, PhD, Danelle Stevens-Watkins, PhD
- August 2013-May 2014 **Shepherds House (drug/alcohol treatment facility), Lexington, KY**
Therapist and Group Leader, Post-Masters Practicum
Site Supervisors: Apryl Tandy, MSW, LCSW, Jason Thomas, MSW, LCSW
Faculty Supervisor: Keisha Love, PhD, Danelle Stevens-Watkins, PhD
- August 2013-April 2013 **University of Kentucky Counseling Psychology Program, Lexington, KY**

Supervisor of Masters Students, Post-Masters Supervision Practicum
Faculty Supervisors: Rory Remer PhD

August 2012-May 2013 **University of Kentucky Counseling Center**
Therapist and Group Co-Leader, Post-Masters Practicum
Site Supervisors: Tina Bryant, PhD, Aesha Tyler, PhD, Susan Mathews, PhD
Faculty Supervisor: Jeff Reese, PhD

February 2012-May 2012 **University of Kentucky Counseling Psychology Program, Lexington, KY**
Supervisor of Masters Students, Post-Masters Supervision Practicum
Faculty Supervisors: Jeff Reese, PhD

January 2012-May 2012 **University of Kentucky Counseling Center**
Interpersonal Process Group Post-Masters Practicum
Site Supervisors: Tina Bryant, PhD
Faculty Supervisor: Jeff Reese, PhD

August 2007-May 2011 **Lifeworks Employee Assistance Program, Minneapolis, MN**
Triage and Financial Counselor, Post-Masters Employment
Site Supervisor: Peter Brower, LCSW

February 2002-May 2004 **Vida para Niños, Guatemala City, Guatemala**
Bilingual Counselor, Pre-Masters Employment [Vida para Niños is a non-governmental organization providing housing and advocacy work in Guatemala for homeless and abused children]
Site Supervisors: Linnea Winkvist, MDiv, Dave Bailey, MA

PUBLICATIONS (PEER-REVIEWED)

Okoli, C. T. C., & **Kodet, J.** (2015). A systematic review of secondhand tobacco smoke exposure and smoking behaviours: Smoking status, susceptibility, initiation, dependence, and cessation. *Addictive Behaviors*, 47, 22-32. doi: 10.1016/j.addbeh.2015.03.018

Okoli, C. T. C., **Kodet, J.**, & Robertson, H. (2015). Behavioral and physiological responses to nicotine patch administration among nonsmokers. *Biological Research for Nursing*. Advance online publication. doi: 10.1177/1099800415579261

Slone, N. C., Reese, R. J., Mathews-Duvall, S., & **Kodet, J.** (2015). Evaluating the efficacy of client feedback in group psychotherapy. *Group Dynamics: Theory, Research, and Practice*. Advance online publication. doi: 10.1037/gdn0000026

Kodet, J. (2012). Social responsibility and prevention in the APA ethical principles of psychologists: Where did they go?. *Prevention in Counseling Psychology: Theory, Research, Practice and Training*, 5, 7-9. Retrieved from www.div17.org/preventionsection/PCP%20Fall%202012%20Issue.pdf

CONFERENCE PRESENTATIONS (PEER-REVIEW NATIONAL)

- Kodet, J.,** Reese, R. J., Duncan, B. L., Bohanske, R. T. (2015, August) Benchmarking Client Feedback Psychotherapy Outcomes of Youth in Poverty Diagnosed With Depression. Poster session presented at the American Psychological Association Annual Convention, Toronto, Canada.
- Lengerich, A. J., **Kodet, J.,** Brown, H. M., Meiller, C., Vasilj, I., Bohanske, R. T., Duncan, B. L., & Reese, R. J. (2015, August) Client Feedback, Outcome, and Recidivism in an Acute Care, Inpatient Setting. Poster session presented at the American Psychological Association Annual Convention, Toronto, Canada.
- Kodet, J.,** Reese, R. J., Duncan, B. L., Bohanske, R. T. (2015, April) Client Feedback Psychotherapy Outcomes of Ethnically Diverse Youth in Poverty: Effectiveness Study of the Partners for Change Outcome Management System (PCOMS). Poster session presented at the Children At Risk Conference at University of Kentucky, Lexington, KY.
- Kodet, J.,** Mason, D. A., Mosely D. V., Cascio, K. A., Simpson, N. B. (2014, August). Experiential workshop addressing privilege, oppression and meritocracy: Pilot study findings. Poster session presented at the American Psychological Association Annual Convention, Washington, D.C.
- Minieri, A. M., & **Kodet, J.** (2014, March). Bringing social justice to the therapy room: Can client feedback help? In R. J. Reese (Chair), *Client Feedback and Counseling Psychology: A Match Made in Evidence-Based Practice Heaven?*. Symposium presented at the 2014 Counseling Psychology Conference, Atlanta, GA.
- Slone, N. C., Reese, R. J., Mathews, S., Hopkins, N., & **Kodet, J.** (2013, October). Evaluating the efficacy of client feedback in group psychotherapy. Poster presented at the Annual meeting of the North American Society for Psychotherapy Research, Memphis, TN.
- Mason, D. A., & **Kodet, J.** (2013, January). Experiential Consciousness Raising Workshop on Privilege and Oppression for Undergraduate Students. Poster session presented at the National Multicultural Conference and Summit (NMCS), Houston, TX.
- Reese, R. J., Toland, M. J., & **Kodet, J.** (2012, August). Validity of a psychotherapy outcome measure: The Outcome Rating Scale. Poster session presented at the American Psychological Association Annual Convention, Orlando, FL.
- Slone, N. C., Reese, R. J., Mathews, S., Hopkins, N., & **Kodet, J.** (2012, August). Using Client Feedback to Evaluate Group Psychotherapy Services: Pilot Findings. Poster presented at the American Psychological Association Annual Convention, Orlando, FL.
- Kodet, J.** (2012, August). Social responsibility in the APA ethical principles of psychologists: Where did it go? Poster session presented at the American Psychological Association Annual Convention, Orlando, FL.

INVITED PRESENTATIONS

Kodet, J. (2014, May). Bringing social justice to the therapy room: Can client feedback help? In R. J. Reese (Chair), *Client Feedback in the UK Research Lab and Counseling Psychology Program*. Invited presentation at bluegrass.org (Community Mental Health Center), Lexington, KY

RESEARCH IN PROGRESS

Kodet, J. (2015). *Effectiveness of psychotherapy of youth in poverty: A benchmarking study of a public behavioral health agency using a client feedback system*. Manuscript in preparation.

Kodet, J., & Mason, D. A. (2015). *Buy Your Opportunity: An experiential workshop addressing economic inequality and meritocracy*. Manuscript submitted for publication.

Lengerich, A. J., **Kodet, J.**, Brown, H. M., Meiller, C., Vasilj, I., Bohanske, R. T., Duncan, B. L., & Reese, R. J. (2015) *Client Feedback, Outcome, and Recidivism in an Acute Care, Inpatient Setting*. Manuscript in preparation.

Toland, M. J., **Kodet, J.**, & Reese, R. J. (2015). *Validity of a psychotherapy outcome measure: The Outcome Rating Scale*. Manuscript in preparation.

PAST RESEARCH EXPERIENCE

November 2011-
July 2014 **Research Assistant, Tobacco Treatment and Prevention Division,
Tobacco Policy
Research Program, University of Kentucky**
PI: Chizimuzo T.C. Okoli, PhD, MPH

August 2011-
December 2011 **Research Assistant, Kentucky Center for Smoke-free Policy, University
of Kentucky**
PI: Ellen Hahn, PhD

May 2010-
July 2011 **Research Assistant, Karen McKinney, EdD, Associate Professor, Bethel
University**
Project: *Making pedagogy political: An examination of simulations on race
analyzed through the lens of critical theory*

TEACHING EXPERIENCE

January 2016 **Guest Lecturer, University of San Diego, Motivational Interviewing,
Restorative Justice and Motivational Interviewing**
Supervisor: Jarrett Ellington, PsyD

February 2014-
June 2015 **Behavioral Preceptor and Teaching Assistant, Introduction to Clinical
Medicine, Behavioral Sciences, University of Kentucky**
Supervisors: John F Wilson, PhD and Raven R. Piercey, PhD

August 2014- November 2014	Teaching Assistant, Patients, Dentists, and Society I, Behavioral Sciences, University of Kentucky Supervisor: Jean Wiese, PhD
August 2014- November 2014	Teaching Assistant, Communication in the Dental Health Care Setting, Behavioral Sciences, University of Kentucky Supervisor: Mitzi Schumacher, PhD
November 2012- March 2013	Workshop Facilitator, Diversity Workshops on Privilege and Oppression Developed curricula and facilitated two separate experiential diversity workshops for undergraduate students.
August 2012	Workshop Facilitator, Ally Development Workshop Developed curricula and facilitated a full day ally development diversity workshop for incoming graduate students in the College of Education

COMMUNITY OUTREACH & CONSULTATION

January 2014- Present	Consultation with Health and Wellness Dept., University of Kentucky Trained staff to implement and routinely utilize brief outcome assessments to evaluate and improve several ongoing wellness programs. I also assisted staff in analyzing their psychological wellbeing outcomes for programs including wellness groups.
July 2015- August 2015	Consultation with Behavioral Sciences Dept., University of Kentucky Assisted professors of Introduction to Clinical Medicine class in developing health disparities unit for first year medical students. Specifically, developed experiential simulations for use in small group settings
November 2013	Diversity Workshop for Hunger and Homelessness Awareness Oxfam Event, Center for Community Outreach, University of Kentucky Developed curricula and facilitated an experiential workshop for 36 undergraduate and graduate students addressing homelessness and hunger within the context of privilege, oppression, social dominance, and meritocracy.

AWARDS and HONORS

August 2014	Recipient, APA Student Travel Award to present research entitled: <i>Experiential workshop addressing privilege, oppression, and meritocracy: Pilot study findings</i> . Division of Counseling Psychology, American Psychological Association
February 2014- June 2015	Teaching Assistantship, Behavioral Sciences, University of Kentucky (\$14,500/year stipend)
August 2011- July 2014	Research Assistantship, Tobacco Treatment and Prevention Division, University of Kentucky (\$14,500/year stipend)

August 2011- Leslie L. Martin Endowed Fellowship, (\$5,000)
May 2012

ADDITIONAL PROFESSIONAL TRAININGS RECEIVED

March 2010 Accredited Financial Counselor (AFC®) designation, Association for Financial
Counseling and Planning Education

